

**Concept of Operations**  
**for the**  
**Access to Archival Databases (AAD) Project**  
**Phase 3 Production**

**NARA IT Project 99-04**



**Version 3.0 (Final)**

**March 22, 2002**

**National Archives and  
Records Administration (NARA)  
8601 Adelphi Road  
College Park, MD 20740-6001**



## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>1</b>
<b>1 INTRODUCTION AND SCOPE OF THE AAD PROJECT .....</b>	<b>6</b>
1.1 INTRODUCTION .....	6
1.2 THE SCOPE OF THE AAD PROJECT .....	7
1.3 BACKGROUND OF THE AAD PROJECT .....	7
1.4 PROJECT RELATIONSHIP TO NARA’S MISSION, STRATEGIC PLAN, AND VISION .....	8
1.5 THE RELATIONSHIP BETWEEN THE AAD PROJECT AND OTHER IT PROJECTS .....	9
1.6 PHASED DEVELOPMENT OF AAD SERVICES .....	9
<b>2 THE CURRENT ACCESSIONED ELECTRONIC RECORDS ENVIRONMENT AT NARA AND ITS IMPACT ON REFERENCE SERVICES .....</b>	<b>11</b>
2.1 THE CURRENT NWME ACCESSIONED RECORDS ENVIRONMENT .....	11
2.2 THE CURRENT NWME REFERENCE SERVICES PROGRAM .....	12
2.3 RESEARCHERS WHO BENEFIT FROM FILE-LEVEL ACCESS .....	14
2.3.1 <i>Providing copies of Electronic Records Files</i> .....	14
2.3.2 <i>Who Orders Copies of Electronic Records Files</i> .....	15
2.4 RESEARCHERS WHO ARE NOT BEING WELL SERVED BY FILE-LEVEL ACCESS .....	16
2.5 OVERALL CONCLUSIONS .....	20
<b>3 JUSTIFICATION FOR, AND NATURE OF, CHANGES WITH IMPLEMENTATION OF THE AAD PROJECT .....</b>	<b>21</b>
3.1 LIMITATIONS IN THE CURRENT OPERATIONAL ENVIRONMENT .....	21
3.1.1 <i>Limitations of the Title List</i> .....	21
3.1.2 <i>Limitations of AERIC as a Search and Retrieval Tool</i> .....	22
3.1.3 <i>Limitations of the MS-ACCESS Database Approach</i> .....	23
3.2 PUBLIC EXPECTATIONS .....	23
<b>4 CONCEPT FOR THE ACCESS TO ARCHIVAL DATABASES PROJECT .....</b>	<b>24</b>
4.1 INTRODUCTION .....	24
4.2 AAD PROJECT RELATIONSHIP TO OTHER NARA IT INITIATIVES .....	24
4.2.1 <i>Relationship to the ERA Program</i> .....	24
4.2.2 <i>Relationship of the AAD Project to the Archival Research Catalog (ARC)</i> .....	25
4.2.3 <i>Relationship to the NARA IT Architecture Initiative</i> .....	25
4.3 PHASES OF THE AAD PROJECT .....	26
4.3.1 <i>AAD Phase 1 - Prototype</i> .....	26
4.3.2 <i>Phase 2 – Pilot - Initial Production Capability</i> .....	27
4.3.3 <i>Phase 3A - AAD Production Capability (before ARC is fully operational)</i> .....	29
4.3.4 <i>Phase 3B - AAD Production Capability (after ARC is fully operational)</i> .....	32
4.4 SUMMARY OF HIGH-LEVEL REQUIREMENTS FOR AAD PROJECT DEVELOPMENT AND INTEGRATION .....	33
4.5 IMPORTANT HIGH LEVEL PHASE 3A TEST AND EVALUATION FACTORS .....	40
<b>5 SUMMARY OF IMPACT .....</b>	<b>41</b>
5.1 PHASE 1 IMPACTS .....	41
5.1.1 <i>NWME</i> .....	41
5.1.2 <i>NH, NPOL</i> .....	41
5.2 PHASE 2 IMPACTS .....	41
5.2.1 <i>NWME</i> .....	42
5.2.2 <i>NWC</i> .....	42
5.2.3 <i>NH, NPOL</i> .....	42
5.3 PHASE 3A AND 3B IMPACTS .....	43
5.3.1 <i>NWME</i> .....	43
5.3.2 <i>NH, NPOL</i> .....	43

## Executive Summary

**Introduction.** This document is Version 3.0 of the Concept of Operations (ConOps) Document for the Access to Archival Databases (AAD) Project. The document reflects the major lessons learned and experiences of the Phase 1 prototyping and the Phase 2 Pilot that have been accomplished to date. This document is being submitted to support the Requirements Review for the development of the AAD Phase 3 production system. Accordingly, this ConOps Document reflects the evolution of the project during the initial requirements definition and the evaluation of the prototype and pilot conducted over the past two (2) years. This version of the ConOps Document defines the current strategy and direction of the AAD Project and is intended to set the stage for engineering development of the Phase 3 Production System consistent with NARA's Managed Evolutionary Development (MED) system life-cycle development methodology. Within this version of the ConOps, the only major change is the introduction of a sub-phase within Phase 3. During Phase 3A, AAD will launch as a stand-alone system, while at the completion of Phase 3B, it will launch from the Archival Research Catalogue (ARC).

**Goal.** The AAD Project responds to, and supports, NARA's mission, strategic plan, and vision to provide NARA's constituents with "*ready access to essential evidence*" at a high level of service. The primary goal of the Access to Archival Databases (AAD) Project continues to be to improve public access via the Internet to accessioned electronic records in structured formats, such as databases, E-mail, and geospatial data. Specifically, AAD is intended to improve the public's access to those structured electronic records that lend themselves to record level access, such as indexes. It is not intended to be a tool to perform sophisticated mathematical analyses on aggregate data, such as is found in most Census databases. AAD is also not intended to provide access to unstructured electronic records, such as the products of standard office automation software. Other major goals and objectives of the AAD Project include:

1. Leveraging the Web-to-database technology inherent in the AAD prototyping that has been performed to date to support public access to two finding aids: the 1930 Census Indices and the Berlin Document Center (BDC) Indices.
2. Developing tools and technology and serve as a testbed for development of architectural components for the *Reference Workbench* under NARA's Electronic Records Archives (ERA) Program.

With the exception of the addition of the Phase 3A/3B change noted above, this Concept of Operations document is essentially the same as the one that was produced for the Phase 2 Pilot dated January 3, 2002.

**Background.** The electronic records collected by NARA over the past quarter century consist primarily of data files and databases. In the current accessioned electronic records management environment at NARA, archival processes for preservation, accession, and reference are neither fully automated, nor fully integrated across the enterprise. At the start of the new millennium, the electronic records being developed and transferred to NARA by other federal agencies are increasing in scope and diversity, and NARA expects substantial sustained growth over the next

five years in its holdings of structured digital data. The content and internal structure of the data files and databases that are being accessioned reflect the broad spectrum of programs and activities of the federal government. In addition, the rapid growth of the Internet is fueling increased public demand for improved online access to all electronic records held by NARA. The AAD Project is designed to help address these basic needs.

**Current Operational Environment.** In the current electronic records environment, NARA has a capability to preserve electronic records through the Archival Preservation System (APS) and a way to verify electronic records through AERIC (Archival Electronic Records Inspection and Control System). Experience with the AAD Phase 1 prototype and Phase 2 pilot to date has, however, identified a number of limitations of the current AERIC metadata model and the data verification process in providing a foundation for on-line access to electronic records. These limitations have been documented in a series of Analysis of Alternatives white papers prepared during Phase 1 of the AAD Project. Perhaps, one of the more significant changes to the Concept of Operations for the AAD Project since Version 1.2 of this document is the acceptance across the AAD Project Team that the metadata used for accessioning and verification purposes is not sufficiently complete for downstream reference and access purposes. When AERIC was developed, while it was hoped that it could be used downstream to support reference, its primary goal was to permit NARA to verify samples of the Agency created metadata. Partly as a result of that lesson learned, one change to the operational concept for the AAD system has been that the AAD system will need to have its own metadata entry, review, and approval capability. This capability was built during Phase 2. Additionally, recent experience in loading, materializing, verifying and searching Agency records using the metadata has demonstrated a need for an *exception handling mechanism* to deal with non-conforming data.

At the current time, the Electronic and Special Media Records Services Division (NWME) still largely addresses the needs of most of its researchers by selling them an electronic copy of an entire dataset. While this level of service may meet the needs of those researchers who need to use an entire database, it does not respond well to the needs of those who seek information that can be answered from one or several records from a database ("record level access"). For these researchers, NWME currently has several options to meet the request, depending largely upon the nature and the timing of the request. These options include:

- Searching MS-ACCESS databases for selected archival series maintained locally by NWME Reference Staff. However, these databases are not directly accessible to members of the public via the Internet.
- For certain records, the NWME staff member may reference a printout of the records and manually conduct a search.
- Lastly, with the introduction of the AAD Phase 2 pilot, if the archival series has been loaded and materialized into the system, the NWME Reference Staff member can search and retrieve records and deliver them to the researcher, either as a printout or as an HTML-tagged Web page attachment in an E-mail. In this regard, the AAD Phase 1 prototype and Phase 2 pilot have been used for limited production support. The Military Prime Contracts File (MPCF) loaded in the AAD system is perhaps the best example to date of the use of the new system to support reference production requests.

In today's environment at NARA, accessioned electronic records are not directly searchable and retrievable by the public across the Internet. Providing improved public access to the National Archives' current and projected accessions of electronic records is the major focus of the AAD Project. As the AAD Project progresses and more and more archival series are loaded and materialized, it is anticipated that the AAD system will become the dominant and preferred method of search and retrieval of selected records. At the current time, metadata is being completed and new archival series are being incorporated into AAD with a goal of having 50 series online with the AAD system when the Phase 3A production system is launched. The AAD Project development approach is to conduct a pilot implementation for Phase 2 of the project with a target for full production with public access via the Internet in Phase 3 (June 2002).

**Phased Development Approach.** Under the AAD Project, NARA will develop, demonstrate, and acquire data access tools and technologies that will provide a single, consistent interface for end-user query and access to structured data with rich, reliable, and flexible search, retrieval, and output capabilities.

The AAD Project is being executed in four (4) phases as follows:

Phase	Identification	Projected Timeframe for Accomplishment
1	Basic AAD Prototype and Proof-of-Concept Demonstration	Delivered May 31, 2001 and remained operational through November 20, 2001. Many architectural components remain part of the current AAD Project Testbed
2	Rollout of AAD Services as a Pilot System with NWME Staff and Other Selected NARA Participants and with Visitors to Archives 2	IOC achieved on November 20, 2001 (currently under evaluation with Version 2.x enhancements being made)
3A	Rollout of AAD Services to the Public as a Standalone Production System via the Internet	IOC planned for April 30, 2001  Actual release to public via Internet planned for June 28, 2002
3B	Rollout of AAD Services to the Public as a Production System via the Internet through ARC	TBD
4	Evolution of AAD Project as part of the ERA Reference Workbench	TBD

**AAD Project Development and Evolution.** Consistent with the direction of the emerging Electronic Records Archives (ERA) architectural vision, the AAD Project will develop and demonstrate feasible and cost-effective services for providing improved public access to structured electronic records. Consistency with the ERA vision includes:

- Access in a uniform manner, independent of the software and/or hardware that the records creator may have used, to as broad a variety of databases and other electronic records exhibiting internal data structures;
- A solution depending as much as possible on commercial off-the-shelf (COTS) products;
- A design that facilitates replacing hardware and software components to take advantage of further improvements in information technology;
- An architecture that supports interoperability with other tools and systems, such as ERA and ARC, through application programming interfaces (APIs) and other techniques that provide the required interoperability without tight binding or coupling between systems and tools; and
- Adherence to NARA's emerging IT Architecture.

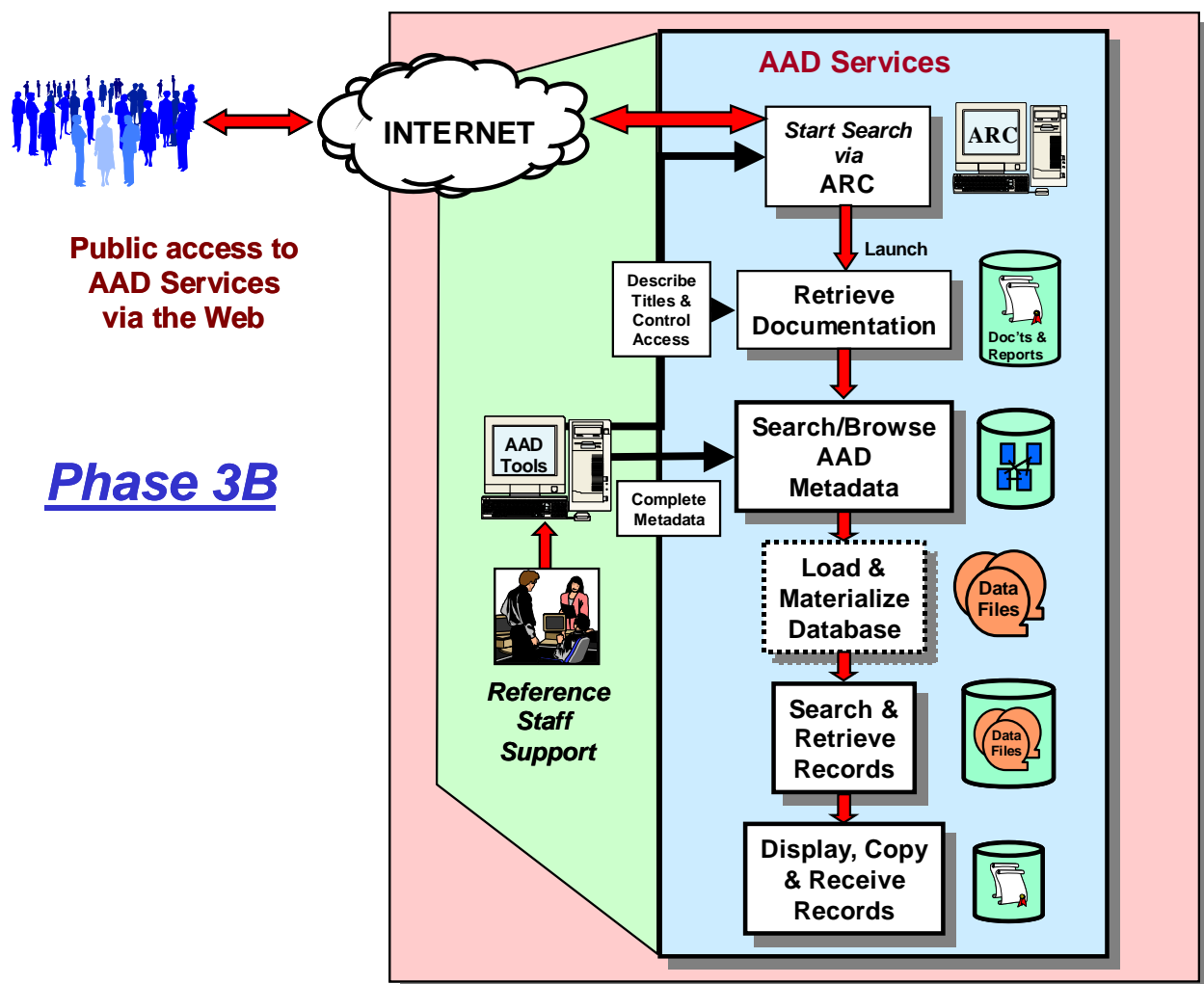
In Phase 1, the AAD Project developed and delivered a working prototype that is currently helping to address immediate problems confronting NARA in improving public access to existing archival databases. For Phase 2, the Phase 1 prototype hardware and software was migrated to become the Phase 2 testbed. The Phase 2 pilot system was implemented in a scalable, high-performance hardware and software environment consistent with NARA's Web architecture. One of the more significant issues currently being addressed in the evaluation of the Phase 2 pilot has been the requirement and process for NARA to complete the metadata to support user-friendly search and retrieval. Within the Phase 1 prototype and Phase 2 pilot systems, sample federal agency data has been loaded and materialized on a virtually daily basis. In this process, a number of issues have cropped up, and these issues are being systematically addressed via Decision Memoranda with the AAD Project Team.

As noted above, the Phase 2 pilot is under active evaluation, and important requirements and systems design parameters are being addressed in such areas as:

- User-friendliness, human factors, and ease-of-use in searching structured databases
- Adequacy, clarity, and usefulness to users of AERIC metadata and federal agency- and NARA-produced documentation (potentially delivered via on-line means)
- Effectiveness of links from NARA descriptions of electronic records; i.e., When researchers get access to the databases, do they find what they had come to expect based on the descriptions?
- Requirements (including the costs and benefits) for completing the AERIC metadata and code lists for high-interest databases
- The requirement for the AAD system to have a new *exception handling mechanism* to support situations where the agency data itself may have errors of a nature where it can not be properly loaded with the specified data type and would consequently adversely impact the search and retrieval process
- Requirements for improved workflow associated with the metadata completion, review, and approval process

- Interface designs with ARC
- Requirements for development of an interim data entry and edit capability for the NARA 1301-compliant, re-engineered Title List to make future ARC entries until ARC is fully operational
- Response times and benchmarks for search and retrieval under certain scenarios and the need to establish quality of service (QoS) performance measures
- Impact (and footprint) of the AAD system on NARA infrastructure systems and networks
- Re-use of AAD Project architectural components and future scalability of the selected approach
- Identification of technical solutions and design requirements to address the Decision Memoranda being developed during the metadata completion process.

Phase 2 has provided an initial production capability that is being used by NWME and selected researchers in Archives 2 research rooms. Phase 2 will rollout AAD services in NWME and will serve as a “stepping stone” for rollout of AAD services to the public via the Internet in Phase 3. When fully integrated with the Archival Research Catalog (ARC) at the completion of Phase 3B, the operational concept for providing AAD services to the public will be as shown below:



# 1 Introduction and Scope of the AAD Project

## 1.1 Introduction

According to ANSI Standard G-043-1992 (“Guide for the Preparation of Operational Concept Documents,” p. 1), the purposes of a Concept of Operations document are to:

- Describe the system characteristics from an operational perspective.
- Facilitate understanding of the overall system goals with the users (including recipients of the products of the system where applicable), buyers, implementers, architects, testers, and managers.
- Form an overall basis for long-range operations planning and provide guidance for development of subsequent system definition documents, such as the system specification and interface specification.
- Describe the user organization and mission from the integrated user/system point-of-view.

In other words, the Concept of Operations document is supposed to describe a proposed system at a very high level to those who will potentially use it and those who are responsible for it. Its principle uses are to facilitate the development of consensus among the system’s principle stakeholders and to guide the development of more detailed documents that will follow, such as the Requirements Analysis and the Analysis of Alternatives. Version 1.2 of the AAD Concept of Operations document accomplished that objective and set the stage for the initial development of Phase 2 and Phase 3 Requirements Documents. These Requirements Documents were backed by a series of Analysis of Alternatives “white papers.”

Consistent with the spirit of NARA’s Managed Evolutionary Development (MED) model, this version of the AAD Concept of Operations “revisits” and largely “validates” the original operational concept for the planned AAD services. As such, this Concept of Operations document provides for a mid-course guidance correction incorporating the major lessons learned during the Phase 2 pilot. This document is intended to serve as the basis for development of the Phase 3A system. Experience with the Phase 2 pilot has shown that effort invested in a good Concept of Operations document can save considerable time and money in the development phases of the project, since the overall direction is crystallized up front.

The Concept of Operations Document for the AAD project is divided into five (5) chapters and two (2) Appendices. The overall plan of the document is as follows:

- **Chapter 1** provides a very broad overview of the AAD project.
- **Chapter 2** describes the current operational environment. It tells how NWME currently provides access to structured electronic data and the limitations of service.
- **Chapter 3** develops a justification for the AAD project, given the limitations of the current environment.
- **Chapter 4** outlines AAD’s high-level functionalities.



- **Chapter 5** shows the operational and organizational impacts of adopting AAD.
- The two **Appendices** provide a table of acronyms and a bibliography.

## **1.2 The Scope of the AAD project**

The primary goal of the AAD Project is to provide public access over the Internet to those structured permanent electronic records in NARA's custody, where access at the record level is meaningful. Within the scope of the project, NARA also seeks to leverage the innate Web-to-database technology in the AAD system to provide improved access to two electronic finding aids. The AAD tools are not currently designed to provide access to relatively unstructured electronic records, such as the products of typical office automation software.

## **1.3 Background of the AAD Project**

NARA is responsible for preserving for posterity historically valuable electronic records produced by all agencies of the federal government and for providing continuing access to these records. In general, the volume and scope of public requests for information about (and contained in) electronic records is growing. In FY 1999, when the AAD Project was initiated, requests for information about or from NARA's electronic records spanned 59 different Record Groups. Public interest in electronic records continues to be high based upon continuing analyses of the number of "hits" on the Electronic and Special Media Records Services Division Web pages. When the 1930 Census is released to the public in April of 2002, public demand for access to the electronic version of the microfilm catalogue will undoubtedly increase. Additional increases are anticipated with the future release of the State Department cable traffic. In large measure, public demand can be expected to increase as NARA makes still more electronic records accessible, through AAD via the Internet. The value of a system such as AAD to the public depends in part on the numbers and types of archival series that are made accessible via the system. The value to the public also hinges on NARA providing convenient and user-friendly access via the Internet.

As stated in NARA's Strategic Plan, "*Users increasingly expect immediate electronic access to information at no cost.*" The growth of Web access and 'e-government,' the availability of electronic access under the Freedom of Information Act, as amended by the Electronic Freedom of Information Act, and provisions of the Government Paperwork Elimination Act will further increase demands for on-line records and services." Meeting the growing demand for public access to electronic records at an acceptable level of service is a major motivating factor underlying the AAD Project. As discussed in Chapter 3, there are a number of technical, operational, and managerial factors that currently limit NARA's ability to respond to the growing public demand for access to accessioned electronic records and non-accessioned electronic records.

The bottom line is that NARA cannot currently provide the public with any access to electronic records except through the purchase of copies of entire physical files. A single physical file may include several million records. To the researcher only interested in one or several records in a dataset, this has become unsatisfactory. The AAD Project is intended to address that

fundamental limitation and growing concern by providing for direct search and retrieval of databases and collections of structured data records by the public online.

#### **1.4 Project Relationship to NARA's Mission, Strategic Plan, and Vision**

**Strategic Plan.** As stated in NARA's Strategic Plan, Ready Access to Essential Evidence: The Strategic Plan of the National Archives and Records Administration, 1997-2007, NARA's mission is to:

“Ensure for the Citizen and the Public Servant, for the President and the Congress and the Courts, *ready access to essential evidence*.”

In meeting this mission, the NARA Strategic Plan sets forth the following four (4) goals:

- Goal 1: Essential evidence will be created, identified, appropriately scheduled, and managed for as long as needed.
- Goal 2: Essential evidence will be easy to access regardless of where it is or where users are for as long as needed.
- Goal 3: All records will be preserved in an appropriate environment for use as long as needed.
- Goal 4: NARA's capabilities for making the changes necessary to realize our vision will continuously expand.

The Access to Archival Databases (AAD) Project is designed to support NARA's mission, strategic plan, and vision to provide NARA's constituents with “*ready access to essential evidence*” at a high level of service. Given the growing importance of the Internet and the increasing impact of electronic records as essential evidence in modern society, the AAD Project is placing increased emphasis on developing timely, user-friendly, search and retrieval services. Collaterally, through the metadata completion effort, the AAD Project is demonstrating the need for increased emphasis on the entire process of creation, management, and use of electronic records to streamline the process.

**Annual Performance Plan.** Additional impetus for the development of AAD services, is discussed in Long-Range Performance Target 3.3 in NARA's *FY-2002 Annual Performance Plan (APP)*, Performance Target 3.3 notes the development of a:

“*Production system to provide interim capability for access to accessioned electronic records in structured form.*”

NARA's Annual Performance Plan (target 2.1) further states that records will be provided to visitors at NARA facilities within one hour of the request. In a modern distributed and networked electronic environment, the very notion of a “*physical visitor*” needs to be challenged. In an Internet environment appropriate to electronic records, NARA's visitors are “*virtual visitors*”, and NARA's domain and “welcome mat” extend to the Internet as a whole, not just to the physical spaces in the Research Rooms in NARA's facilities. This vision of ubiquitous public access to unrestricted electronic records is completely consistent with NARA's

Strategic Plan, which states in Goal 2 that: *“Essential evidence will be easy to access regardless of where it is or where the users are located”*. Additionally, the concept is firmly supported with the recent establishment of the Electronic Records Archives (ERA) Project, under which the AAD Project is being designed and developed.

### **1.5 The Relationship Between the AAD Project and Other IT Projects**

A major key to success is that the AAD Project be consistent with, and evolve along the lines of, NARA’s emerging electronic records architectural concept, referred to as ERA (Electronic Records Archives). The Archivist established the ERA program in NARA Notice 2000-074 as a major new initiative to address key technical issues in the creation, management, and use of electronic records. The plan and future direction for the ERA are now heavily discussed in NARA’s Strategic Plan (revised in 2000). The Archivist’s establishment of the ERA program reflects the observation that *“...every type of record we manage – including textual, cartographic, and audiovisual – already exists in digital form.”* In NARA Notice 2000-074, the Archivist also stated that the *“Planning for ERA, then, must be broad and thorough.”* Key components of the ERA vision are:

- ❑ Overcome technological obsolescence in a way that preserves demonstrably authentic records.
- ❑ Build a dynamic solution that incorporates the expectation of continuing change in information technology and in the record it produces.
- ❑ Find ways to take advantage of continuing progress in information technology in order to maintain and improve both performance and customer service.

The AAD Project will conform to the ERA vision and serve as a testbed for future ERA development. To the maximum extent practicable, the AAD system will be designed to use commercial products with a view towards migration into the full ERA system when that is developed.

NARA’s primary method for reference will be for the users to access the Archival Research Catalog (ARC) as the principal “finding aid” for identifying relevant records series and archival holdings. After locating records series of interest in ARC, users will be passed seamlessly to the AAD system to browse and search the actual data in the structured data files that are available in AAD. We expect this functionality to be available to the public when Phase 3B of AAD has been completed.

### **1.6 Phased Development of AAD Services**

Consistent with the direction of the emerging ERA architectural vision, the AAD Project will develop and demonstrate feasible and cost-effective strategies for providing improved public access to structured electronic records. The AAD Project is being executed in four (4) phases as follows:

1. Phase 1 – Basic AAD Prototype and Proof-of-Concept Demonstration
2. Phase 2 – Rollout of AAD Services as a Pilot System with NWME Staff and Other Selected NARA Participants and with Visitors to Archives 2
3. Phase 3A – Rollout of AAD Services to the Public as a Standalone Production System via the Internet
4. Phase 3B – Rollout of AAD Services to the Public as a Production System via the Internet fully integrated with ARC
5. Phase 4 – Evolution of AAD Project as part of the ERA *Reference Workbench*

Details of the functionalities that will be developed in each of the phases are provided in Chapter 4.

## **2 The Current Accessioned Electronic Records Environment at NARA and Its Impact on Reference Services**

The primary mission of the AAD Project is to develop and demonstrate capabilities for improved public access to accessioned databases and other structured records (like E-mail), either currently in the custody of the National Archives or expected to be received from federal agencies. This chapter therefore describes the *current* environment in which the Electronic and Special Media Records Services Division (NWME) provides reference services for accessioned electronic records.

### **2.1 The Current NWME Accessioned Records Environment**

NARA's custodial program for electronic records, now more than 30 years old, originated well before the era of the personal computer or the Internet. Most of the series of electronic records that have been appraised as permanent, and most of the electronic records transfers to NARA have been databases or data files. Since data processing applications continue to support federal government programs, NARA anticipates that federal agencies will continue to transfer valuable electronic records in databases in a wide variety of formats. Providing a high level of service to historically-valuable electronic records over the long-term and the steps to achieving that goal have always been a central feature of NARA's electronic records program.

NWME accessioning archivists currently use the Archival Electronic Records Inspection and Control System (AERIC) utility to verify datasets when they are received from federal agencies. Verifying an accession means determining to what extent the records transferred to NARA correspond to what the agency of origin said it would transfer. The verification is accomplished by keying some or all of the fields in the electronic records accession into AERIC and having the utility tell the archivist whether the records conform to the parameters that the agency provided in its documentation. The metadata that are a by-product of AERIC verification constitute digital documentation that subsequently can be used in a reference system, such as the AAD system.

As a result of the AAD Phase 2 pilot, we now know that the metadata developed during the AERIC accessioning and verification process, while a valuable starting point, is not necessarily the best (or sufficient) metadata to support user-friendly, downstream reference, access, and search and retrieval functions for public users. The limitations in the AERIC metadata model and the verification process itself (with perhaps an over-reliance on the use of VARCHAR as a data type) have been widely discussed across the AAD Project Team and are documented in three Analysis of Alternatives white papers prepared during Phase 1 of the AAD Project. These white papers include:

- *Alternatives for Development of a Common, Enterprise-Wide, Metadata Server*
- *Alternatives for Handling Complex Queries with Metadata (e.g., parent-child, zoned decimals, grouped data fields, Soundex, date/time)*
- *Metrics and Benchmarks for Completion of the Metadata*

## 2.2 The Current NWME Reference Services Program

**Finding Aids: Title List.** The reference process sometimes begins with the *Title List*. The *Title List* is currently the primary finding aid for accessioned electronic records. The *Title List* is maintained as a DB-2 relational database on a mainframe computer at the National Institutes of Health (NIH) campus in Bethesda, MD. It is available to the public through NARA's web site. Recently, the AAD Project converted the *Title List* data to a true-relational database compliant with NARA Directive 1301. In Phase 2, the AAD Project also developed a Web-based 1301 data entry capability call the Records Description Tool (RDT). NWME is using the RDT to perform data entry and edit capability to create and edit descriptions of series and files that will be made available in AAD that will be suitable for migration into ARC. The method of migration is still under consideration.

At the current time, a public version of the "old" *Title List* is still available via the Internet in two (2) versions: as an extracted ASCII file for download using FTP and also in HTML on the NARA homepage. The full *Title List* database is not directly accessible to the public, but was used in the translation and migration to the RDT. The *Title List* database includes entries for over 14,000 files; but there are over 100,000 electronic records files in NARA's holdings. Once researchers locate information of interest in the *Title List*, they typically contact NWME through traditional means (e.g., telephone, mail, E-mail, etc.). Because the *Title List* is only a subset of all data files in the National Archives, it is considered preliminary and partial. New entries and updates to the *Title List* required for AAD purposes will be made using the new RDT data entry and edit capability developed under the AAD Project. These entries will ultimately be used to populate the ARC database.

**File-Level Access.** With few exceptions, access to NARA's current accessioned electronic records is still largely limited to offering researchers, on a cost-recovery basis, copies of full files of electronic records on a variety of electronic media and with a range of technical specifications. This situation is expected to change rapidly with the introduction of AAD services in Phase 3. Major metadata completion and data loading activity is currently underway within NWME. Part of this activity includes identifying agency- and NARA-produced documentation and scanning, and converting it for ultimate on-line public access as part of the AAD services. Additional efforts are underway to add archivist's notes to the metadata to provide assistance and guidance to public users.

In the current environment, paper copies of documentation packages for the files are offered on a cost-recovery basis. Traditionally, this basic form of access responds to expectations from researchers who want or need to use a file(s) of NARA's accessioned electronic records for some analytical purpose on computer hardware and software available to them. In other words, their interest is to use the predominant kind of electronic records that federal agencies thus far have transferred to NARA, for their own data analysis purposes. These researchers, rarely, if ever, search for and retrieve specific records from the electronic files. Some may, in the course of their analysis, create new files that are subsets of records with common characteristics in order to analyze them as a distinct universe. For example, from the files of electronic records on all World War II prisoners of war (POW) and civilian internees, a researcher might create a subsetted file with records for all personnel who were incarcerated at Stalag Luft 1 Barth-

Vogelsang, Prussia, and then analyze them in an effort to discover whether they all came from the same military organization, or were captured in the same place, etc. Such traditional use of agency documentation and agency data files is far different than that of the researcher seeking to retrieve his or her father's POW record. AAD will enable this class of researchers to make preliminary determinations about whether a specified data file or database actually contains sufficient data of interest, in a form that suits their needs. It will also provide additional empirical information, beyond what is available in technical documentation, useful in planning for using the data prior to making a purchase. While these possibilities may not have been important for many of the databases that NARA has accessioned up to now, they are likely to become more important given the size and complexity of collections expected in the future.

**Current Record-Level Access to Accessioned Electronic Records.** The AAD system will address the needs of the researcher who wants very specific records from a data set, like the individual who wants to know if his father was a POW, or a graduate student who wants all State Department cable messages concerning Henry Kissinger. This more specific level of inquiry might be termed "record-level access," as opposed to the "file-level" access that was described above. By way of background to the AAD Project, NWME has taken a number of steps to enhance record-level access to electronic records in the past, as described below:

- *State-level extract casualty lists.* In 1998, NARA posted on its homepage, state-level casualty list reports extracted from the electronic records of casualties from the Korean and Vietnam conflicts. They can be viewed or downloaded at <http://www.nara.gov/nara/electronic/korvnsta.html>.
- *Ad-hoc query using AERIC.* In a very limited manner, NWME staff have used (and continue to use) AERIC metadata and the MS Query utility within AERIC to search, on behalf of researchers, for specific records responsive to their query(ies), retrieving and printing any records found.
- *MS-ACCESS Utility.* Further, to meet some of the demand for access to specific records in electronic databases, NWME staff have developed for in-house staff use, a desktop application for search and retrieval from a small selection of frequently requested electronic records files, including records of World War II prisoners of war and civilian internees; and casualty records from the Korean and Vietnam conflicts.
- *Other.* As a general rule, NWME does not routinely produce extracts, tabulations, or formatted printouts for any of its electronic records holdings except two (2) casualty files from the Korean and Vietnam Wars, mentioned above. In addition, in some cases where specific records or data elements within electronic records are restricted, NWME may make an extract of the records in a disclosure-free version, also known as a public use file.

The principle features to note in the current NWME reference services program are that:

- The current basic mode of access to NARA's accessioned electronic records is predominantly file-level, rather than record-level access;

- With the exception of the state casualty extracts lists on the NARA home page, the research public does not currently have on-line access to any of NARA's accessioned electronic records;
- NWME reference staff use of AERIC for record-level access is limited because it was never intended to support that function directly, and using AERIC for access interferes with verification productivity. In addition, metadata produced for electronic records verification may not include full field information for a file or series of files and often does not include all code tables. Other limitations in the metadata model are addressed in the series of Analysis of Alternatives white papers mentioned above.
- The use of existing MS-ACCESS databases is not scalable to support enterprise reference applications and the requirements for on-line public access via the Internet.

## **2.3 Researchers Who Benefit from File-Level Access**

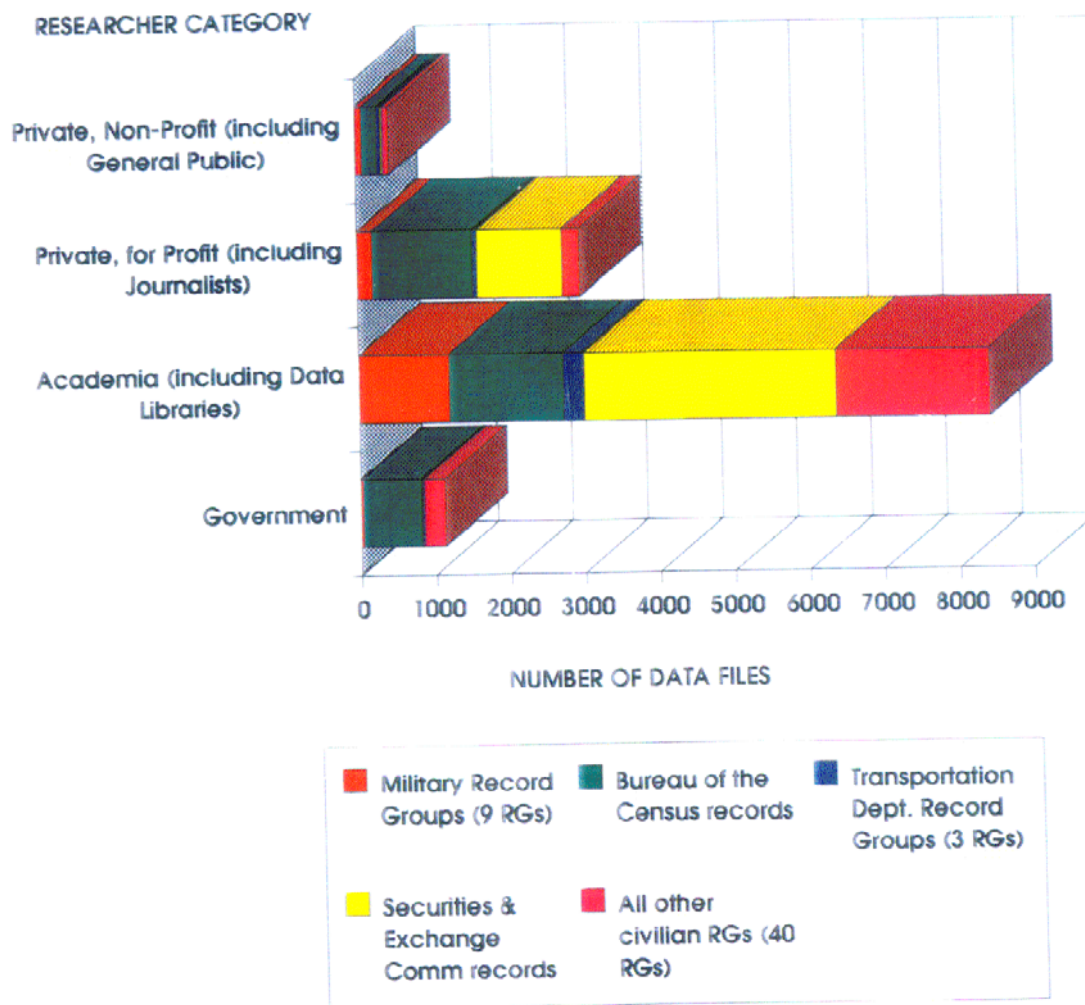
### **2.3.1 Providing copies of Electronic Records Files**

Given the character of the databases that NARA has accessioned and until the ubiquity of the Internet and personal computing, the dominant expectation for access to electronic records in databases was for file-level access. NARA responded with a program that offers copies of an electronic records file(s) on standard computer-readable media, like a magnetic tape, cartridge, and more recently CD-ROM, on a cost-recovery basis. Researchers can then analyze the records with computing hardware and software appropriate for their work and available to them, either personally or organizationally. This is the mode of access for electronic records that NARA has offered for the past three decades.

Currently, NWME provides access to hard copy versions of agency created documentation for electronic records files on-site in the Electronic Records research room, or through photocopied reproductions on a cost-recovery basis. Most researchers opt to order their own copies of documentation packages, and do not personally come to NARA to review electronic records documentation. Over the years, newer technologies have meant that NARA, keeping pace with these changes, has provided copies of the data in new forms of output media. At present NARA provides copies of electronic records files on diskettes, CD-R, 3480-class cartridge, or 9-track tape, with a choice of technical specifications to meet customers' needs. NARA currently is exploring further enhancement of its access options to include transferring copies of files to researchers on-line, using the capability known as FTP, or file transfer protocol.



### 2.3.2 Who Orders Copies of Electronic Records Files



**Figure 2-1. Data Files Copied by Researcher Category, FY 1987 to FY 1997 (Number of Files Shipped = 13,004)**

As indicated in Figure 2-1, over the 11-year period from FY 1987 to FY 1997, NWME (and its predecessor organizations) provided copies of 13,004 files for an average of 1,182 files per year. In FY 1999, NARA shipped nearly twice the average, or over 2200 electronic records files to researchers. During the first half of FY 2000, NARA shipped copies of more than 1800 electronic records files to researchers. In the late '90s, the requests for file reproductions grew rapidly. More recent data is beginning to indicate a leveling of such requests. While there has been some leveling of the requests for file-level access, it is instructive to look at the relative numbers and types of researchers that have requested such access, including what they have requested.

The files sent in FY 1999 to academic, government, media, private sector and lay public researchers throughout the U.S. and abroad represented historic electronic records from 25 different federal agencies or organizations (Record Groups), and in the first half of FY 2000, from 26 Record Groups. Offering researchers access by providing copies of electronic records files that they can keep and use in an unlimited manner, with their own computing hardware and software, is still a popular mode of access to NARA's electronic records holdings.

Figure 2-1 is provided primarily to show the historical distribution of NWME's customers and the Record Groups of the files that they have typically been requesting. It shows that historically, the overwhelming number of files copied were from its most advanced users, specifically from academics and "private, for profit" users. In general, "Academic" and "Private, for Profit" researchers can be viewed as relatively sophisticated users who need copies of electronic records files for a variety of data analysis purposes. They tend to be using computer-based tools to analyze data in ways intended to produce new knowledge, using the primary documentation that NARA's electronic records represent as their source material. They undertake this analysis to inform their teaching or as the basis for publication. An example would be economists who analyze annual files from the records of the Department of Transportation, measuring several hundred megabytes of data each, documenting airline ticket dollar prices and schedules in an attempt to assess the impact of airline deregulation on the pricing structure of the airline industry. "Private, for Profit" researchers (including Journalists) typically have directed and focused goals and objectives for their analysis. Additionally, they bring a profit motive with regard to their projects. Users primarily interested in file-level access can benefit from AAD by being able to sample the records before deciding which files they wish to acquire.

As also shown in Figure 2-1, "Private, Non-profit" researchers (including the General Public) are currently the smallest class of researcher identified in the statistics on files ordered. This class of users is viewed as a major target for development of user-friendly search and retrieval services as they are more likely to be seeking record-level access within electronic records files. Sophistication of this class of users can vary widely from highly sophisticated and IT-literate personnel on a LAN/WAN supporting a private non-profit organization to relatively unsophisticated home users with dial-up Internet access.

## ***2.4 Researchers Who Are Not Being Well Served by File-Level Access***

While we anticipate continuing demand for reproductions of electronic records files, we also know that file-level access does not meet the needs of all researchers. It is unresponsive to most requests for information or records "from" our holdings. For example, a citizen seeking to document his internment as a civilian during World War II to qualify for compensation, probably has a once-in-a-lifetime need to access his own electronic record in the database that has thousands of similar records. He needs that record reproduced, probably as some kind of a report, with all items of information that were stored electronically in code interpreted with their literal meaning. He is not interested in, nor does he need, a copy of the entire file with the records of all his fellow internees. By contrast, a copy of all the records of the file in electronic form is a valuable primary source ripe for iterative analysis by the historian or sociologist that is

trying to discover common characteristics of all the civilian internees in World War II. Currently, NWME research staff respond to individual requests for POW records by using the MS-ACCESS utility NWME staff has developed for this purpose. If record(s) is (are) found in the search, they print out the individual record(s) in a customized report form and send by surface mail. NWME provides this service on the typical 10-day turnaround time that they apply to all research letters.

Table 2-1 below, prepared at the start of the AAD Project, summarizes the numbers of reference requests that came to NWME during FY1999, excluding formal orders. While this data is a bit dated, it is still considered representative of the public demand for records-level access. Indeed, a number of the Record Groups (and series) in Table 2-1 are currently being entered and loaded into the AAD Phase 2 pilot system with a goal of being accessible for production support with the Phase 3A rollout.

NWME codes each request as either: "about records," "from records," or "other." Generally, staff code requests like "Do you have any information/records on...." as "about records," and requests like "How many contracts were there between the Boeing Corp and the Department of Defense in 1969" as "from records," and requests for price quotations for copies of electronic records files and documentation among the requests shown below as "other." The number of inquiries for information "from" electronic records (the column from Table 2-1 that is shaded) is particularly pertinent to the development of the AAD Project, because these inquiries necessitate search and retrieval of particular records similar to the process we plan to automate and make "user-friendly" with the AAD Project.

**Table 2-1. Identification of the Types of Inquiries Received from the Public Regarding Electronic Records As a Function of Record Group, FY1999**

<b>Record Group</b>	<b>Description</b>	<b>Total Inquiries (via E-mail)</b>	<b>Inquiries for Information "About" Records (via E-mail)</b>	<b>Inquiries for Information "From" Records (via E-mail)</b>	<b>Other, Including Requests for Order information</b>
15	Veterans Administration	23 (16)	5 (3)	17 (12)	1 (1)
25	National Labor Relations Board	16 (5)	3 (0)	2 (2)	11 (3)
29	Bureau of the Census	254 (72)	52 (19)	19 (6)	183 (47)
38	Chief of Naval Operations	9 (8)	1 (0)	5 (5)	3 (3)
47	Social Security Administration	11 (7)	7 (5)	0 (0)	4 (2)
58	Internal Revenue Service	22 (3)	6 (2)	1 (0)	18 (1)
195	Federal Home Loan Bank Board	24 (0)	7 (0)	4 (0)	13 (0)
197	Civil Aeronautics Board	14 (5)	7 (1)	0 (0)	7 (4)

<b>Record Group</b>	<b>Description</b>	<b>Total Inquiries (via E-mail)</b>	<b>Inquiries for Information “About” Records (via E-mail)</b>	<b>Inquiries for Information “From” Records (via E-mail)</b>	<b>Other, Including Requests for Order information</b>
218	U.S. Joint Chiefs of Staff (relating to Vietnam)	17 (12)	6 (5)	6 (3)	5 (4)
220	Various Committees, Commissions and Boards	6 (4)	4 (2)	0 (0)	2 (2)
266	Securities and Exchange Commission	90 (20)	17 (5)	3 (0)	70 (15)
306	U.S. Information Agency	44 (32)	14 (10)	3 (1)	27 (21)
307	National Science Foundation	6 (4)	2 (1)	2 (1)	2 (2)
319	Army Staff	16 (14)	6 (6)	7 (5)	3 (3)
330	Office of the Secretary of Defense (primarily contracts and Korean and Vietnam casualty records)	804 (554)	217 (155)	516 (376)	71 (23)
349	DOD, Joint Commands (all relating to Vietnam)	13 (10)	7 (5)	4 (3)	2 (2)
389	Office of the Provost Marshal General (WW II POW data)	104 (70)	21 (12)	74 (55)	45 (3)
398	Department of Transportation	35 (6)	8 (0)	0 (0)	27 (6)
407	Adjutant Generals Office, 1917- (Korea and Vietnam Army casualties)	616 (402)	132 (84)	445 (309)	39 (9)
443	National Institutes of Health	28 (4)	5 (1)	2 (0)	21 (3)
467	DOT, Research and Special Programs Administration	21 (7)	5 (3)	0 (0)	18 (4)
479	Merit Systems Protection Board	11 (6)	3 (2)	1 (1)	7 (3)
483	Office of Thrift Supervision	13 (0)	6 (0)	1 (0)	17 (0)
512	Health Resources and Services Administration	10 (5)	7 (4)	1 (0)	2 (1)
888	Donated Historical Materials	26 (18)	4 (3)	4 (2)	18 (13)

<b>Record Group</b>	<b>Description</b>	<b>Total Inquiries (via E-mail)</b>	<b>Inquiries for Information “About” Records (via E-mail)</b>	<b>Inquiries for Information “From” Records (via E-mail)</b>	<b>Other, Including Requests for Order information</b>
All other	Sum of 34 other Record Groups (RGs) where the total number of inquiries for a given RG is $\leq 5$	70 (39)	35 (27)	8 (5)	27 (7)
999	Inquiries that are not RG specific	983 (676)	461 (324)	364 (277)	158 (75)
<b>Totals</b>	<b>Totals</b>	<b>3286 (1999)</b>	<b>1048 (676)</b>	<b>1489 (1063)</b>	<b>749 (260)</b>

Note: Record Group “999” represents inquiries that are not specific to any particular record group, or that are unrelated to any electronic records files in NWME custody. Record Group “888” indicates donated historical materials.

In reference to Table 2-1, several observations are germane to the AAD project. Nearly half (1489 of 3286) of the current numbers of inquiries that NWME receives (those listed as inquiries for information “from” records) are potential candidates for online AAD search and retrieval services. AAD is intended to address the needs of these researchers, since it would permit them to view, manipulate, and retrieve specific records on-line, instead of having NWME reference staff perform the search and retrieval. The other interesting figure to note is that about half (1063/1999) of those who requested data “from” the records sent their request to NWME by E-mail. This may indicate a comfort level on the part of these researchers to transacting their research business over the Internet.

When comparing the data in Figure 2-1 with that in Table 2-1, it is evident that the Record Groups represented by the bulk of the files ordered are quite different from those in which researchers wanted information from the records. Although the data in these two illustrations come from different times, the experience filling orders in FY 1999, the year represented by the inquiries in Table 2-1, was similar to that of the eleven-year period, FY 1987-1997 represented in Figure 2-1. In particular, orders for full files from the Records of the Bureau of the Census (RG 29), and of the Securities and Exchange Commission (RG 266) represented 44 percent of all files ordered in FY 1999. Orders from Donated Historical Materials (shown as RG 888), accounted for another 30 percent of all files ordered. Inquiries related to these three groups, which together accounted for 74 percent of all full files ordered, account for less than two percent of the Table 2-1 “inquiries for information from records” column. Put another way, the NWME experience to date is that in general, the electronic records in demand at the file-level are different files than those for which there is significant public expectation for record-level access. With respect to the types of electronic records that have predominated in transfers to the National Archives, AAD is intended to respond to the access needs and interests of researchers who are not currently well served by the file-level access option, and whose interest in record-level access is generally from different kinds of files than those most in demand by data analysts seeking file-level access.

## **2.5 Overall Conclusions**

Continued technological innovation has produced new computing capabilities, rising information access expectations, and a broadening of the spectrum of people who can utilize computers and potentially, electronic records. Yet, in lieu of the development of AAD services, NARA currently is able to provide record-level access to only a tiny portion of its electronic records files that is analogous to bringing a box of paper records to a researcher, and allowing him/her to search through the papers in the box until he/she finds relevant material. What the above analysis indicates is that generally speaking, requests for specific records like the individual POW record discussed above come from a wholly different group of researchers than those who seek file-level access. We expect requests for specific records to accelerate as NARA begins to accession new forms of electronic records, like diplomatic cable records or E-mail records. We also anticipate continuing demand for copies of full files of electronic records.

NARA, like most agencies of the federal government, is experiencing pressing public demand for on-line access to our holdings, coupled with expectations for searching and retrieving specific records from accessioned electronic files. There are critical differences between NARA and most agencies in that NARA must provide access to the records it accessions forever and that the electronic records for which NARA must offer record-level access come, potentially, from every agency in the federal government. The records were designed to meet the programmatic needs of the agencies that created them. The primary challenge is to develop a generic access mechanism for these highly heterogeneous records and to design this mechanism so that it is relatively immune from technological obsolescence and that it can take advantage of continuing improvements in relevant technology. This is the access challenge AAD is attempting to meet. The AAD Project will develop a system that will be a general-purpose, computer access utility for searching and retrieving specifically identifiable records in NARA's holdings of structured electronic records in databases.

### **3 Justification for, and Nature of, Changes with Implementation of the AAD Project**

This section summarizes the justification for developing the AAD system. The AAD Project's goal is to permit researchers to gain access through the Internet to specific records within NARA's historic electronic structured records, such as databases and electronic mail systems. Currently, as discussed in Chapter 2, researchers can acquire a copy of NARA's electronic records files, on a cost recovery basis. While this level of service meets the needs of those who want copies of entire files, it does not begin to address the needs of people who need only specific records from within the accessioned electronic files.

#### **3.1 Limitations in the Current Operational Environment**

As discussed in Chapter 2, currently most researchers access NARA's historic datasets by ordering copies of the agency documentation and complete files containing the electronic records. The files are sent on electronic media, and the documentation is normally sent as paper copies. Other researchers can only search and retrieve accessioned electronic records by consulting with members of the NWME Reference Staff who, together with the researcher, must first identify appropriate Record Groups and then conduct the appropriate searches among the datasets. This can be a time-consuming process that sometimes takes days to accomplish.

*In the current accessioned electronic records environment, however, there is no system specifically designed for search and retrieval of the records.* This is a fundamental operational limitation. Designing, developing, and implementing such a capability for NARA to support the public represents the major thrust and focus of the AAD Project. NWME Staff has been able to search and retrieve specific records using native capabilities in its accessioning system (e.g., AERIC – Archival Electronic Records Inspection and Control System) and has developed some specialized and local MS-ACCESS databases to enable search and retrieval of individual records in five electronic records series. There are, however, significant limitations in NWME's current ability to meet the public's growing requirements for record-level access. The major limitations of the *Title List*, and the native AERIC and the MS-ACCESS-based approaches are briefly described below.

##### **3.1.1 Limitations of the Title List**

In the current operational environment, the primary finding aid for accessioned electronic records is the *Title List*, a public version of which is accessible as an ASCII file via the Internet. However, as noted earlier, only about 14,000 of the 100,000 datasets are currently listed in the *Title List*, limiting its utility. The public version of the *Title List* (as a finding aid) is not directly searchable in the same sense as the keyword search capability in, for example, the NAIL catalog database. The HTML version of the public *Title List* has been indexed on NARA's Web site. Thus, a researcher can enter a "text string" in NARA's main search window, and if the *Title List* contains the string, the researcher will get a "hit." However, this type of search is largely a "hit

and miss” proposition. The researcher may well get hundreds of other hits across the NARA Web site. Additionally, the titles of the records series provided within the *Title List* tend to be a bit arcane. The titles are not descriptive in the same sense as a more detailed abstract, and the titles (in the list) frequently contain acronyms that may not be immediately apparent to the researcher. Notwithstanding these limitations, the *Title List*, itself can be downloaded into a word processor, where the titles can then be searched via a simple “find” command native to the word processor. The *Title List* can also be browsed via a Web browser on NARA’s Web site with word search then limited to the “find” command for the displayed Web page. The acronyms used in the *Title List*, however, are not explained in the browser window.

The *Title List* also does not contain a listing of the data fields that are contained in the files, so it is generally not possible to determine whether the files in a series contain any data fields of interest without reviewing the documentation or consulting with the Reference Staff member, who must then consult with the documentation and/or access the AERIC metadata. As noted previously, the Title List does not provide a more descriptive abstract of the records series. More fundamentally, however, once a researcher has located the materials that he/she wants, the automated portion of the process ends. At this point, the researcher who seeks record level access must contact NWME to request a search of the relevant files. The researcher who wants a copy of the file(s) requests a price quotation to order a complete copy.

### 3.1.2 Limitations of AERIC as a Search and Retrieval Tool

NWME Reference Staff have used a native *ad hoc* query capability in the baseline AERIC system (Archival Electronic Records Inspection and Control System). AERIC is NARA’s primary system to verify electronic records as part of the accessioning process. The metadata database that is produced using AERIC in the accessioning process has only been in existence since 1993 and is not complete. Another reason that the metadata database is incomplete is that it is not necessary for the verification process for all of the fields to be entered into AERIC and frequently they have not been. The absence of any metadata in AERIC for many accessions, and the incomplete state of the metadata for others restricts even the limited possibility of running *ad hoc* queries on AERIC for reference. The databases, themselves, can also contain additional complex data elements that are not adequately described by the metadata.

Perhaps the most significant limitation is that AERIC is designed for accessioning, not reference. It cannot handle the public demand for search services. When AERIC is used for search and retrieval, the data must also be loaded and materialized. Within AERIC, the data is not physically maintained in an on-line or near-line environment and must therefore be loaded directly from tape cartridges. Finally, searches using AERIC and the specialized MS-ACCESS databases can only be performed in NWME workspaces, and the service is therefore limited to normal business hours.

Another limiting factor in using AERIC for search and retrieval is that accessioned electronic records are contained in physical files stored on archival magnetic tape cartridges produced by the Archival Preservation System (APS). In general, to conduct a search of specific records in response to a reference request, NWME staff must consult the APS database to identify and locate the relevant tape(s) and file(s). The tapes must be manually retrieved, mounted, and



positioned. In addition, the files may span multiple tape volumes. Each file must be copied to disc and loaded as a database in AERIC. Only then can NWME staff execute a query in response to a researcher request. Formulating the query may require staff to consult manual codebooks to identify the specific codes that correspond to the data the researcher wants. This is a time-consuming process which is inherently inefficient given that neither APS nor AERIC were designed to support reference queries and that both systems are heavily taxed just to support the preservation and accessioning functions, respectively.

### 3.1.3 Limitations of the MS-ACCESS Database Approach

As mentioned earlier, in the case of five (5) military series, NWME responded to the demand for specific information from these records with a very different approach. In addition to offering researchers copies of the data sets, NWME has downloaded the raw data and added the code values for coded data elements in the records of these five series into several databases that reference staff has developed in MS-ACCESS 97. The advantage of this approach is that staff can readily search these MS-ACCESS databases and provide a hard copy report to the researcher that the researcher can readily understand. The limitations, however, are that the database is available to the staff of NWME only, not to the research public, and that a special *ad hoc* effort was necessary to create a MS-ACCESS database for each file. Even if such a utility were made publicly available, it is neither very user-friendly nor sufficiently scalable to meet the rather large need for data materialization and retrieval at an acceptable level of service.

## 3.2 Public Expectations

While future customer demand for access to electronic documents and data via the Internet is anticipated to be “very high”, it is not known with any degree of quantitative accuracy “how high it can or will be.” Every time that NARA has offered a new electronic capability, however, demand has quickly soared to outstrip the capability and to require still more service.

There is an extremely important operational concept at issue relative to timely public access to accessioned electronic records in a modern Internet environment. The Internet is intrinsically a “*self serve*” information search and retrieval environment. The NARA Annual Performance Plan also states that records will be provided to visitors at NARA facilities within one hour of the request. In a modern distributed and networked electronic environment, the very notion of a “*visitor*” probably needs to be challenged. In an Internet environment appropriate to accessioned electronic records, potential customers have no need to visit a NARA facility. Access can be provided much more readily and probably more efficiently via the Internet.

The agency’s strategic plan, plus the severe limitations of the current environment, coupled with growing public demand for improved on-line access, are strong motivating factors that help define the scope and purpose of the AAD Project. NARA’s mission and vision state that NARA will provide *ready access to essential evidence* and easy access to the evidence *regardless of where it is or where users are for as long as needed*.

## 4 Concept for the Access to Archival Databases Project

### 4.1 Introduction

This chapter discusses the relationship of this project to other IT projects, as well as to existing systems, identifies a phased approach to building AAD through operational scenarios, and describes the high-level requirements of AAD.

### 4.2 AAD Project Relationship to other NARA IT Initiatives

#### 4.2.1 Relationship to the ERA Program

Figure 4-1 depicts the preliminary NARA vision or conceptual model for the Electronic Records Archives (ERA). At the current time, it must be emphasized that the architectural vision for ERA is still preliminary and subject to revision and refinement. The ERA concept anticipates development of capabilities to support accessioning, maintenance of the electronic archives, and reference. These capabilities will interface with other activities relating to the Records Schedules and the Archival Research Catalog (ARC).

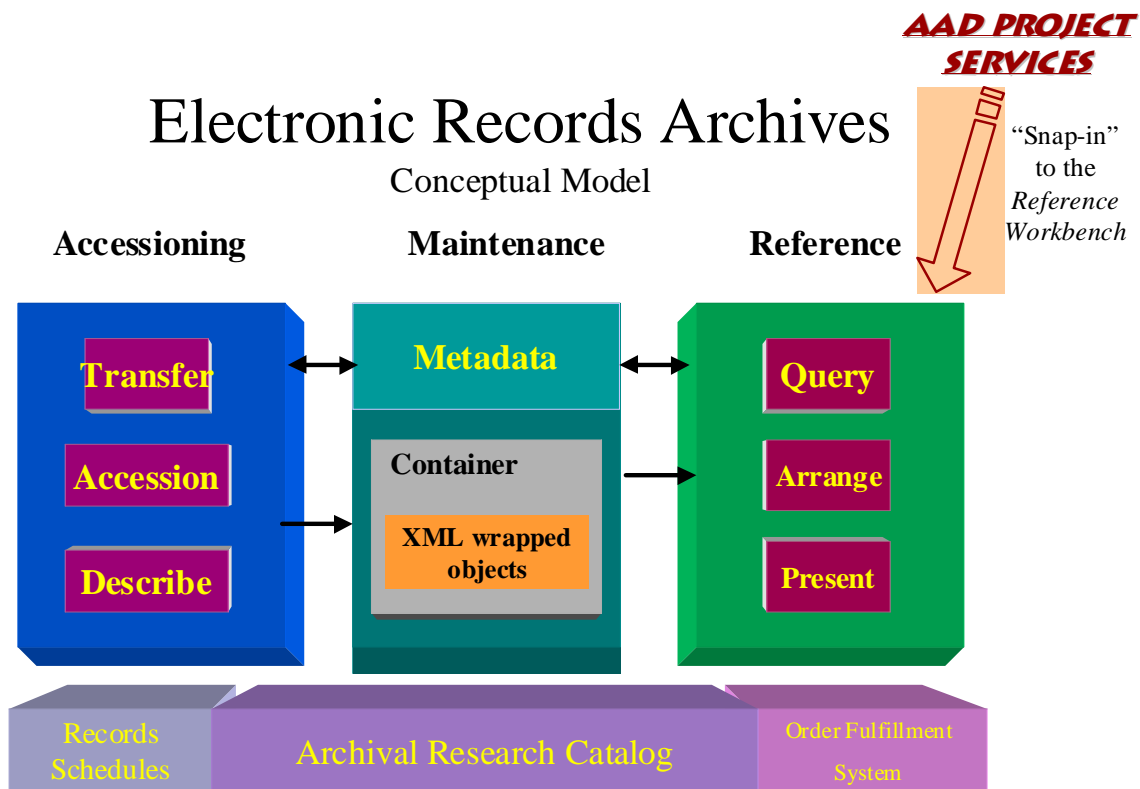


Figure 4-1 – Preliminary NARA Vision of the Electronic Records Archives

Figure 4-1 also shows that the access utility developed in the AAD Project will fit into the ERA *Reference Workbench*. Specifically, this utility will constitute a ‘snap-in tool set’ for query, arrangement and presentation of databases and similar collections. As indicated in the AAD Project description in Chapter 1, the AAD Project will develop and demonstrate reusable IT architectural components including systems, software, and tools (collectively - “services”) that will evolve to become part of the *Reference Workbench*.

The scope of the query and reference tools anticipated to ultimately become a part of the *Reference Workbench* is not fully defined, thus future evolution of the AAD Project can best be viewed as only covering a “part-of” the *Reference Workbench* concept. In the preliminary ERA concept, the *Reference Workbench* is also envisioned to “unwrap” the electronic records that are retrieved from the Records Repository, and to arrange them for presentation to the user. In Phase 2, the AAD Project demonstrated some of the required unwrapping, arrangement, and presentation capability for structured databases including sample State Department cable traffic and US Trade Representative E-mail indices that were tagged in XML.

#### 4.2.2 Relationship of the AAD Project to the Archival Research Catalog (ARC)

As shown in Figure 4-1, in NARA’s future electronic records environment, the primary starting point for reference will be for the users to access the Archival Research Catalog (ARC) as the principal “finding aid” for identifying relevant records series and archival holdings. Both ARC and AAD are currently in development. Efforts are underway to establish an appropriate applications program interface (API) between ARC and AAD, which will be operational at the completion of Phase 3B development effort for AAD.

In operation, it is anticipated that the user will use ARC to identify relevant records series and/or file units to be searched. ARC will then handoff the user and the user’s request to search particular records series and/or file units to the AAD server for more comprehensive and more detailed searching of the metadata for that series and/or file unit. Users will then be able to retrieve archival records via AAD services, FTP, and physical media – as appropriate. The method of retrieval is dependent on a number of factors: the volume of material to be retrieved, the speed of the user’s network connection, the urgency of the request, the cost to complete the order, public demand, and time of day. These will be balanced against the limits of AAD storage and processing capabilities.

#### 4.2.3 Relationship to the NARA IT Architecture Initiative

The Information Technology Management Reform Act (ITMRA) requires all federal agencies to develop a comprehensive Information Technology Architecture (ITA). NARA is currently in the process of developing its overall enterprise information technology architecture (ITA) including establishing systems engineering processes and procedures, technical standards and guidelines (TSGs), and high-level systems architectural development and integration principles (for IT Project Managers). As part of that effort, NH defined a systems development approach entitled Managed Evolutionary Development (MED). The AAD Project is being implemented using the

MED methodology. NH also recently developed an IT Architecture system diagramming methodology that is intended to communicate client/server architectures unambiguously between systems architects and developers.

The AAD Project will be compliant with the emerging NARA standards-based IT Architecture. In that regard, the AAD Project Team will work closely with NH staff members to help develop the enterprise IT Architecture and to ensure that the AAD Project is fully compliant with the architecture.

### 4.3 Phases of the AAD Project

This section describes the three (3) phases of the AAD project through the use of operational scenarios. During Phase 4, AAD will become part of the ERA Reference Workbench. This is a functional view of the system presented from a user's point of view.

#### 4.3.1 AAD Phase 1 - Prototype

Figure 4-2 depicts the overall concept of operations that was implemented during the course of development of the Phase 1 prototype. Requests for records from the Military Prime Contracts File (MPCF) are perhaps the best example to date where the prototype has been used to support production (on a trial basis).

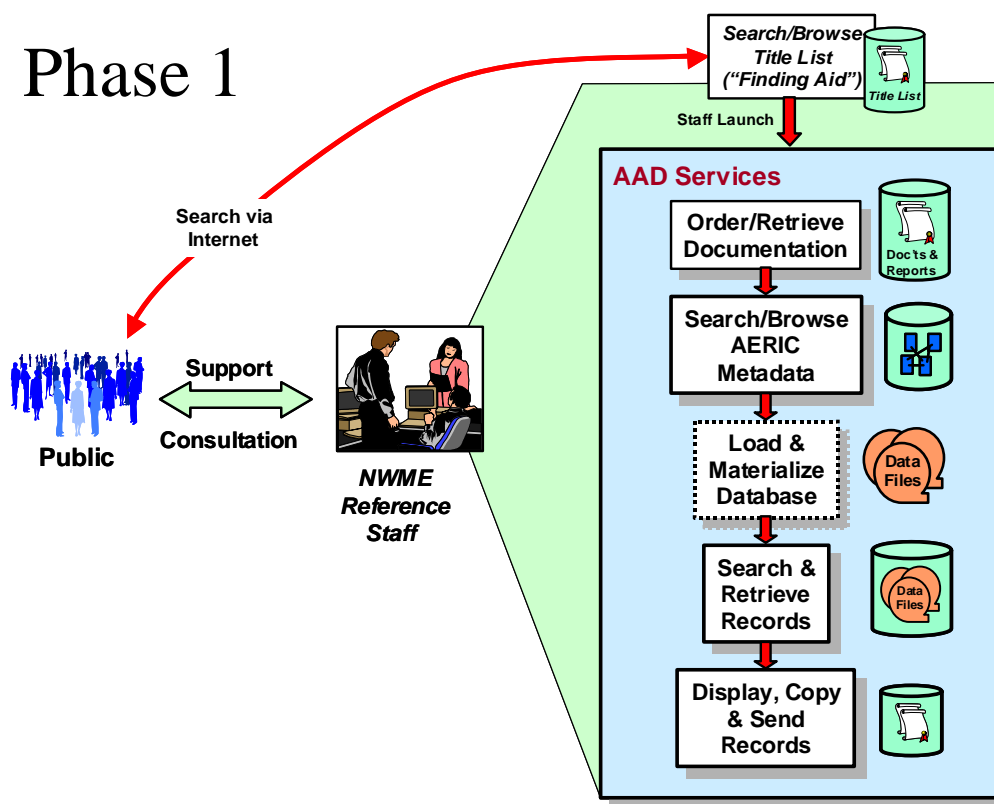


Figure 4-2: Operational Concept for the Development of Phase 1

The operational concept for Phase 1 was implemented primarily via COTS products with the major AAD applications software components. A full description of it can be found in the deliverable entitled, "System Description and Data model for the AAD Project, Phase 1 Prototype," July 16, 2001.

For those datasets that were part of the Phase 1 prototype, the following customer support scenario was supported as shown in Figure 4-2:

- Members of the public were able to browse the existing *Title List*, and/or consult with the Reference Staff member to find records series of interest. NAIL was used for specific records series where the catalog existed for the records of interest.
- Members of the public were able to communicate with the Reference Staff member via telephonic means, E-mail, regular mail, and/or facsimile and discuss the nature of the access request.
- Members of the public were able to order and receive documentation. For documentation that has been scanned and converted, and where the communications data rate is acceptable, the documentation may be sent to the researcher via electronic means. It may also be sent on a CD-ROM and/or photocopied.
- Members of the public were able to consult with the Reference Staff member on specific requirements for the search. In the background, the Reference Staff member can use the AAD system to search AERIC metadata to determine whether there are any data elements that may be of interest to the researcher. The Reference Staff member will assess the completeness of the AERIC metadata (e.g., data layouts and code lists) and can advise the customer accordingly.
- Assuming that the customer wants to proceed with the query of the databases that are accessible in the Phase 1 prototype, the Reference Staff member need to ensure that the databases are loaded and materialized. The staff members were able to then conduct the required search(es) using the AAD client/server environment.
- The Reference Staff members were able to then advise the customer on the status of the request and any search and retrieval results. Costs were then assessed and adjudicated, as required. The AAD system was able to retrieve the electronic records (e.g., results of the query – in an appropriate electronic format (e.g., HTML) for display and transmission back to the customer).
- The Reference Staff members were able to make arrangements to send the query results back to the customer. Fees were able to be charged and collected, as necessary, on a cost-recovery basis. If feasible, network connections and bandwidth permitting, the Reference Staff members were able to send back the query results as a file attachment to an E-mail message. Alternately, the results may be sent as printed material.

#### 4.3.2 Phase 2 – Pilot - Initial Production Capability

During Phase 2, the AAD project developed an initial production capability for AAD client services operable at Archives 2. The Phase 2 initial operational capability (IOC) was realized in

late November 2001. During Phase 2, the development contractor (SAIC) has functioned as an Application Service Provider (ASP) under a Service Level Agreement (SLA).

With the advent of Phase 2 capability, the most significant change in the customer support scenario is that AAD client services are now available across NARANET in Archives 2. Another change is that the number of series available through AAD increased for the 6 completed during the prototype to 17 when Phase 2 became operational. The Phase 2 system is currently in the O&M stage with over 30 archival series and over 170 files now loaded. The on-site reference staff will be functioning as proxy users and testers of the system.

Phase 2 has provided NWME with practical experience in providing AAD services in a distributed on-line environment. NWME continues to support external customers from outside a NARA facility, the process being similar to the Phase 1 scenario described above, except that NWME reference archivists now have more datasets available to search and browse with AAD than they had in Phase 1. In Phase 2, in lieu of a direct interface to ARC, the AAD services are launched from a List of Titles (re-engineered to be NARA Directive 1301-compliant) to be maintained within the system. Figure 4-4 depicts the high-level customer business support functions and processes for the initial production capability that have been developed as part of Phase 2.

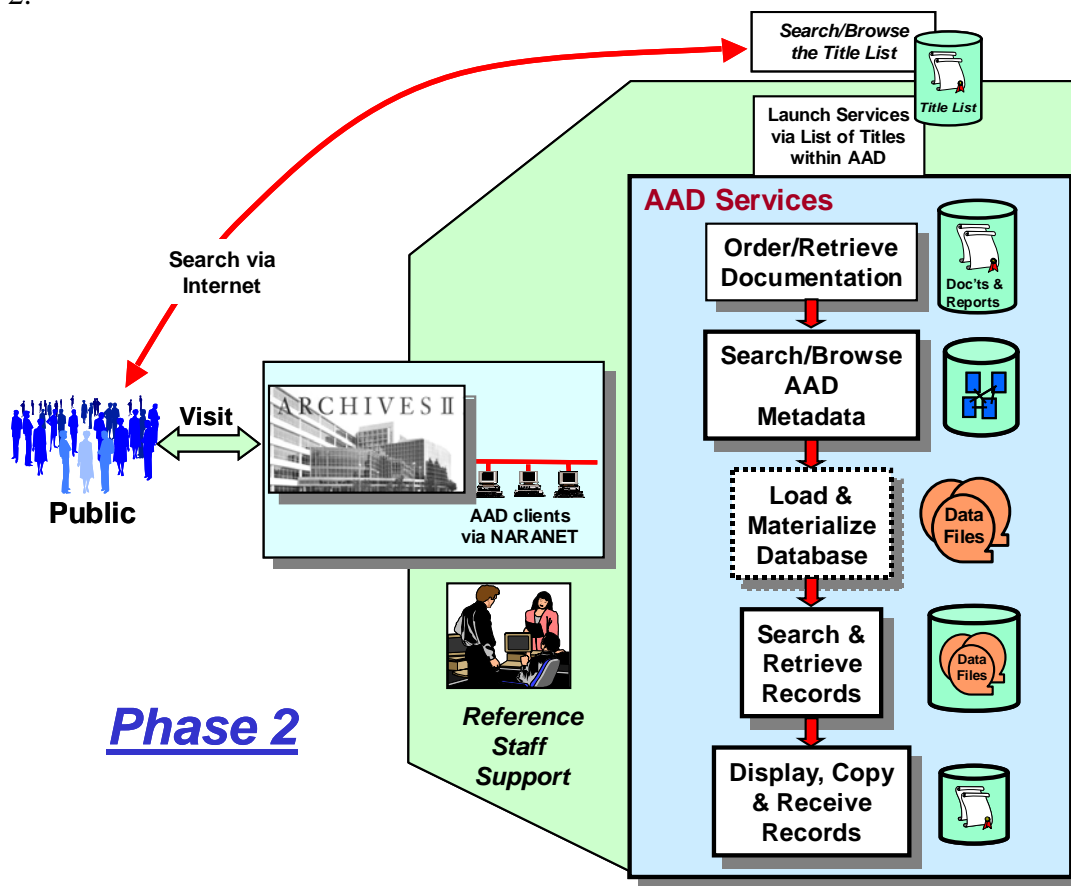


Figure 4-4: High-Level Customer Business Support Functions for Phase 2

For those datasets that were part of Phase 2, the following customer support scenario includes the following:

- NARA staff are able to support visitors to Archives 2 to provide access the AAD services for some 30 series via workstations within Archives 2. Visitors are able to browse the List of Titles within AAD, and/or consult with the Reference Staff member to find records series of interest. NAIL can also be used for specific records series where the catalog exists for the records of interest. Note: In Phase 2 operations, there are currently some limitations to public researcher access to the AAD system even within Archives 2, since no archival series has yet been declared to be fully ready for public use. Thus, any visitors currently need to coordinate with NWME to gain any access to the system. Also, the system remains under UserID and password controls.
- Researchers are able to receive documentation via electronic means when loaded in the system. For documentation that has been scanned and converted, and where the communications data rate is acceptable given the volume of material, the documentation may be sent to the researcher via NARANET.
- Researchers are able to consult with the NWME Reference Staff member on specific requirements for the search. The researcher is able to directly access the complete metadata and determine whether there are any data elements of interest. With support from the NARA reference staff member, the researcher can use the AAD client to be able to pose and formulate specific queries.
- With the limitations above, while at Archives 2, researchers are able to use the AAD services to query the database and to then display, and/or copy the results.

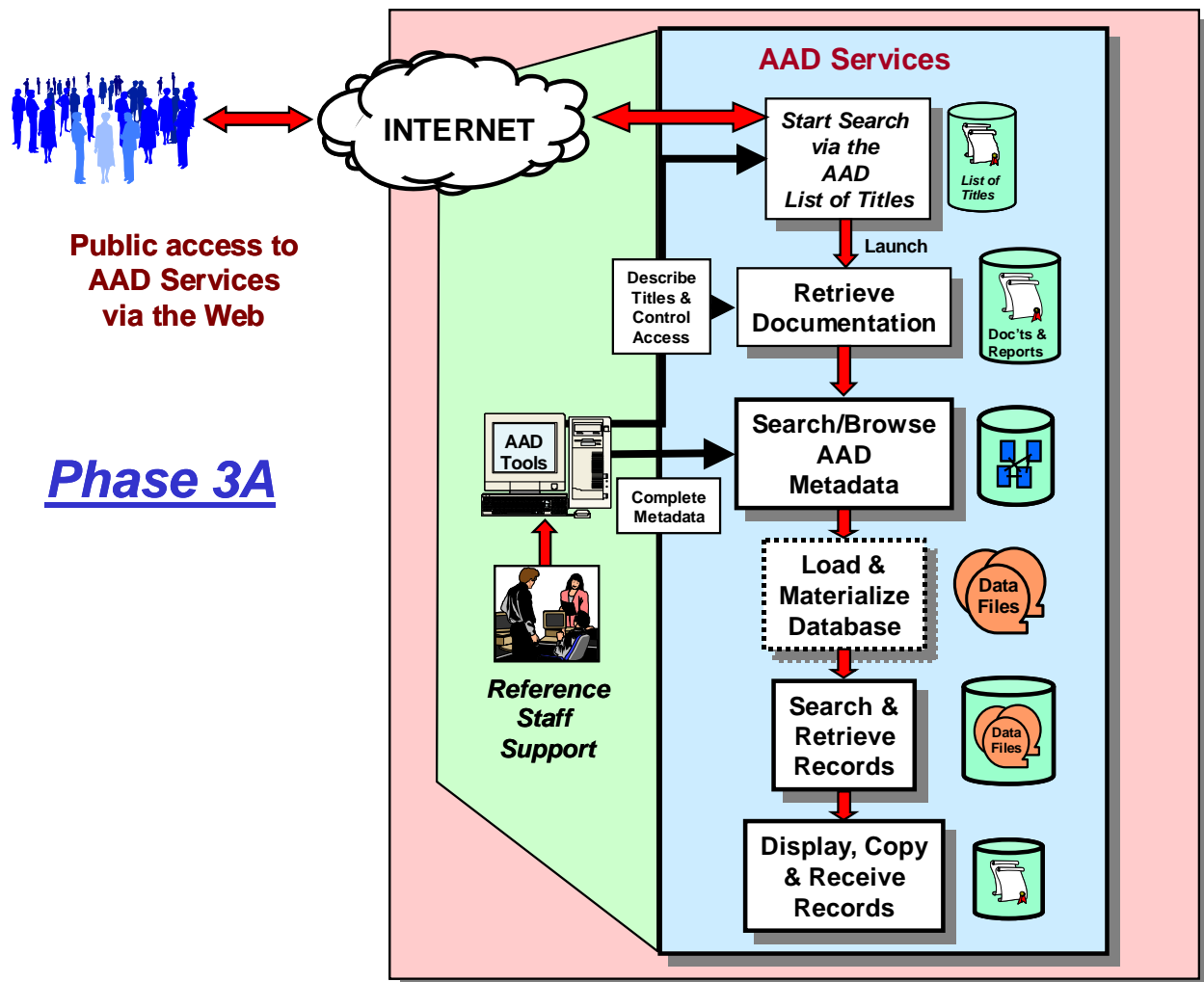
While the pilot includes limited production capability, the principle purpose of the Phase 2 system was to help NARA refine its requirements for Phase 3A. Besides the performance of the pilot system, NARA was able to evaluate the types and extent of customer support that is needed for effective use of the AAD system. In Phase 2, a separate testbed has been maintained as a development and testing platform. Also, in Phase 2, the system supported continued prototyping and development of Web-to-database technology for access to finding aids, such as the BDC Indices and the 1930 Census Indices.

#### 4.3.3 Phase 3A - AAD Production Capability (before ARC is fully operational)

Phase 3A will develop a production capability for AAD client services to be operable to the public via the Internet. At this juncture, it is important to note that all of the potential customer support requirements and impact are not fully known. The most significant change in the customer support scenario will be that AAD client services will be available to the public via a thin Web client across the Internet. The system will provide capabilities for search and retrieval of accessioned electronic records, using the completed metadata to facilitate search and retrieval of the records.

In Phase 3A, quality of service will be an important design criterion impacting the concept of operations and customer acceptance. High public interest databases will be pre-loaded and materialized. Performance of the system will be monitored and tracked. The AAD client/server

architecture will provide the user with an estimate of time and costs needed to fulfill user requests. Figure 4-5 depicts the high-level customer business support functions and processes for the working prototype to developed and evaluated as part of Phase 3A.



**Figure 4-5. High-Level Customer Business Support Functions in Phase 3A**

From a customer support perspective, the high-level steps in the Phase 3A process for the datasets that are part of Phase 3A will include:

- NARA reference staff will use tools within AAD to complete the metadata and to make descriptive data entries for series/files entered into the system. The RDT will be used to control which series and files are made accessible to the public via the Internet.
- NARA customers will be able to access the AAD services related to 50 series of electronic records via workstations and PCs on the Internet.
- NARA staff and the public will be able to access the AAD server configuration database to determine which databases are readily accessible.



- NARA staff and the public will be able to launch the AAD services client via a variety of methods.
- NARA staff and the public will be able to retrieve NARA- and Agency-produced documentation for series and files of interest.
- NARA staff and the public will use the AAD services to search and/or browse the completed metadata to determine the utility of the database(s).
- NARA staff and the public will be able to search and retrieve specific records in electronic files through AAD.
- The user will then be able to use the AAD client to be able to pose and formulate specific queries. The system will provide the query results appropriately formatted for display in a Web browser window.
- AAD will alert the customer in advance when processing a request either is likely to exceed allowed parameters of time, storage or other AAD resources, or when it would require payment.

**Anticipated Public Demand in Phase 3A.** As the AAD System approaches its release to the public as an on-line Web browser-based database query application in Phase 3A, a significant amount of consideration must be given to achieving an acceptable level of performance and to providing the scalability necessary to support public demand. Scalability is an important design feature for most enterprise applications. It will be especially important for the AAD system. In Phase 3A, it is anticipated that the AAD system will demand a very high volume of use and wide array of users will be accessing its services 24 hours a day, 7 days a week, and from all over the world. Such a high volume of users has the ability to access a tremendous amount of data, thus placing a strain on AAD system resources.

Unlike traditional client/server applications, which have a limited number of users, the number of users served by the AAD system is virtually unlimited. In Phase 3A, when AAD is initially deployed to the public, it may expect a few thousand hits per month. That number can easily turn into hundreds of thousands or even millions virtually overnight especially with the “flash crowd” phenomenon as discussed in the O&M Analysis of Alternatives white paper. Additionally, one can readily envision growth in public demand for AAD services with something as simple as potentially large numbers of high school students using the convenient, Web-based and official records access capability of NARA’s AAD services to develop their civics term papers (including the night before it is due!) Other researchers and students will undoubtedly use the system for science fair projects in the area of sociology. Other members of the public can be expected to use the system, if for no other reason than to satisfy curiosity and to perhaps “learn something”. In that regard, federal agency E-mail and the State Department cable traffic are expected to be very popular. The bottom line is that the potential of the AAD system must not be underestimated during the design phase. The challenge will be, of course, to design an Internet-based access system that is scalable and provides acceptable levels of performance at a reasonable cost.

For Phase 3A, the AAD system must be built to support an increasing number of users and, therefore, must be scalable. This does not mean that the AAD system has to handle 500,000 hits per day upon its initial release to the public. It simply means that the AAD system needs to be designed with scalability in mind so that when certain thresholds are reached, adding additional

hardware to the system configuration can maintain an acceptable level of performance as user volumes increase. This will probably be the most important aspect of the AAD system's application design for Phase 3A.

#### 4.3.4 Phase 3B - AAD Production Capability (after ARC is fully operational)

Phase 3B will continue basic Phase 3A AAD production capability for client services to be operable to the public via the Internet. The major difference between Phase 3A and Phase 3B is that during Phase 3B, AAD services will be launched from ARC, rather than from a list of AAD titles, and we anticipate being able to provide service on approximately 70 series total. Phase 3B may also provide for an interface of AAD services with the Order Fulfillment and Accounting System (OFAS). This future interface with OFAS to handle high-volume copying and delivery of records is currently under active consideration.

Figure 4-6 depicts the high-level customer business support functions and processes as of the completion of the Phase 3B development effort.

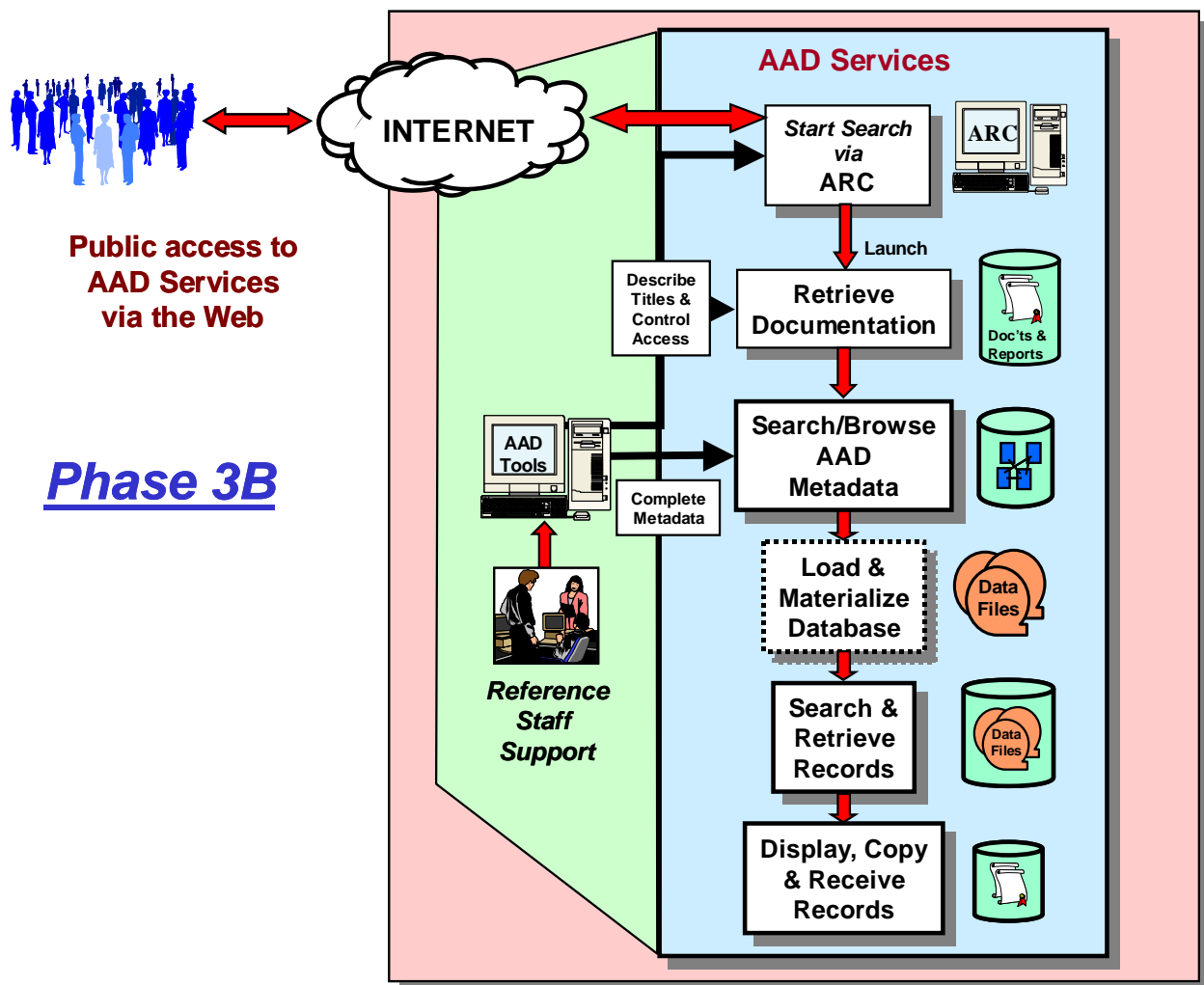


Figure 4-6. High-Level Customer Business Support Functions in Phase 3B

From a customer support perspective, the high-level steps in the Phase 3B process for the datasets that are part of Phase 3B will include:

- NARA customers will be able to access the AAD services related to 70 series via workstations and PCs on the Internet. Once the interface between ARC and AAD has been defined and fully implemented and NWME has the ability to create and edit its ARC entries at the completion of the Phase 3B development effort, customers will use ARC as the primary finding aid for accessioned electronic records.
- They will be able to launch the AAD services client via a variety of methods.
- They will use the AAD services to search and/or browse the completed metadata to determine the utility of the database(s).
- They will be able to search and retrieve specific records in electronic files through AAD.
- They will be able to access documentation via predominantly electronic means.
- They will be able to access the AAD server configuration database to determine which databases are readily accessible.
- The user will then be able to use the AAD client to be able to pose and formulate specific queries. The system will provide the query results appropriately formatted for display in a Web browser window.
- AAD will alert the customer in advance when processing a request either is likely to exceed allowed parameters of time, storage or other AAD resources, or when it would require payment.

#### **4.4 Summary of High-Level Requirements for AAD Project Development and Integration**

The following briefly summarizes the high-level AAD Project development and integration requirements:

- **Overall.** The AAD Project shall provide a single, consistent interface for end-user query and access to structured data with rich, reliable, and flexible search, retrieval, output, and visualization capabilities. Consistent with the initial prototyping and development activity accomplished in Phase 1 and Phase 2, AAD services shall be developed to be responsive to quality of service (QoS) requirements and shall be scaleable to anticipated future high public demand for access via the Internet. The AAD services shall be consistent with NARA's mission, Strategic Plan and "Vision" statement, and the ERA concept.
- **ERA Compatibility.** Throughout all phases, the AAD Project shall be forward compatible with the emerging NARA ERA concept, especially the "*Reference Workbench*."
- **Compliance with NARA's IT Architecture and Continuing Requirements Definition.** The AAD Project will be developed using the Managed Evolutionary Development (MED) methodology, as defined in the NARA's *Systems Development*

*Life-Cycle (SDLC)* document. The Phase 2 pilot system has been developed in accordance with NARA's target IT Architecture including processes, procedures, and documentation standards. Additional AAD concept and requirements development is also needed to address unique aspects of providing access to selected non-accessioned records. The Phase 2 pilot system was initially operational on November 20, 2001. The detailed requirements for Phase 3A and 3B will be refined based upon NARA experience with the Phase 2 pilot system. At the current time, Phase 3A of the AAD services is assumed to be initially operational in April 2002 and available to the public at the end of June 2002.

- **Testbed.** A testbed, which was an outgrowth of the Phase 1 prototype, will be maintained across the life-cycle of the AAD Project to serve as a vehicle for continued enhancement and test and evaluation of AAD services. The major goals and objectives of a testbed are defined as part of the NARA MED methodology. The reader is referred to the Testbed Analysis of Alternatives white paper for additional details on the scope and purpose of the testbed for the AAD Project.
- **Interface to the Archival Research Catalog (ARC).** ARC is NARA's primary finding aid for its archival materials and at the completion of the Phase 3B development, researchers will launch AAD from ARC. Until this integration with ARC can be put in place, the AAD system will be launched from the list of available file titles maintained within AAD. During Phase 2, the development contractor (e.g., SAIC) worked with NWME and the ARC Project Team to re-engineer the current *Title List* data to be NARA 1301-compliant in anticipation of making it available to the public via AAD. SAIC has also built a data entry and edit tool, the Records Description Tool (RDT), to provide NWME with a capability to create and edit descriptions of series and files that will be made available in AAD.
- **Basic Search and Retrieval Functionality.** The AAD Project shall provide robust search and retrieval services. The detailed requirements are discussed in the Requirements Documents for Phase 3A. In the spirit of the MED model, the Phase 3A Requirements Document was derived from the practical experience gained during the development and evaluation of the Phase 2 pilot. Additional details will be provided in the AAD system design documentation.
- **Archival Data Files Included in AAD.** During Phase 3A, the AAD system's primary purpose is to provide the public with access to accessioned archival data files that are most appropriate to records level access via the Internet. During FY2001, NWME and the AAD contractor identified numerous series of electronic records for eventual public access through AAD. NWME plans to have 50 series ready for public access at the completion of Phase 3A and a total of 70 series at the completion of Phase 3B. As part of this process, the AAD team has learned a great deal about the nature and limitations of historic datafiles. NWME has had to balance public access requirements with issues of records integrity. Another high-level requirement that NWME has identified is the development of an automated workflow process to more

efficiently manage the completion of the metadata and the preparation of data files for AAD.

- **Support for Non-Accessioned Records.** The primary mission of the AAD system is to provide access to accessioned archival databases. However, during Phase 2 NARA elected to use AAD technologies to provide access to two electronic finding aids, Berlin Document Center and the 1930 Census Catalogue. In providing access to both accessioned electronic records and finding aids, the AAD system developed an appropriate user interface that distinguishes between these classes of materials, including (but not limited to) providing the user with proper links to the appropriate NARA business unit for reference support. For selected electronic finding aids, the AAD system will provide a mechanism to create and/or edit such materials under strict user profile controls.
- **Use of Metadata in the AAD Environment.** Based upon experience with the Phase 2 pilot, the AAD system will maintain its own copy or version of the APS and AERIC metadata databases for accessioned electronic records, and their equivalent for the finding aid databases made accessible via AAD services. The data derived from APS also includes technical specifications for files that are copied and transferred to the AAD system. With the AAD system maintaining its own copy or version of the APS and AERIC metadata databases, the system will need to maintain strict version control on any changes made to the metadata to facilitate traceability back to the original source of the metadata. The decision to develop a common metadata server may also be deferred to Phase 4 where the AAD services are integrated into ERA (Electronic Records Archives).

Consistent with the direction and guidance of the NARA AAD Project Team, the current rationale for permitting AAD to maintain its own variant of the metadata databases is that the metadata entered for accessioning and verification purposes within AERIC and for preservation with APS is not necessarily the most suitable or complete metadata needed for reference, search, and public access purposes. A similar comment can also be made regarding metadata associated with NARA-produced records and other non-accessioned records.

For non-accessioned electronic records that will be accessible via AAD services, metadata analogous to that in AERIC and APS will also need to be developed and maintained within the AAD environment.

- **Agency- and NARA-Produced Documentation.** The AAD Project shall be able to deliver selected agency- and NARA-produced documentation in an electronic format via on-line methods. Should on-line delivery of documentation not be practicable, the AAD Project shall provide for the user to order documentation via manual means. With the introduction of AAD services, there will be a continuing requirement for the user to have access to agency- and NARA-produced documentation. Currently, the documentation exists largely on paper. A typical set of agency documentation may be upwards of 300 pages. The pages may also be of low optical quality for scanning

purposes. Phase 1 demonstrated the use of raster scanning and optical character recognition (OCR) as possible vehicles for delivery of the documentation and reports. This activity was documented in an Analysis of Alternatives white paper. As a practical matter, a raster-scanned set of 300 pages of documentation is nominally 9 megabytes in size. Depending on the speed of the networking connection, delivery of such a large file may not be practicable. As a matter of experimentation and future development, the AAD Project will continue to consider various alternatives for conversion of documents and reports to an intelligent electronic form and for flexible delivery to the user via on-line and manual methods.

- **Projected Public Demand and 24/7 Operations.** At the current time, projected public demand for AAD services is not known to any degree of accuracy; however, the projected public demand is assumed to be qualitatively “high” especially if the AAD services provide access to State Department cable traffic. Also, it is generally not possible to project the growth of public demand, insofar as the introduction of convenient, user-friendly AAD services may snowball and stimulate even more public demand for more Internet access to more electronic archival holdings. The AAD Project Team recognizes that the quantitative uncertainty in projected public demand has an impact on defining the requirements for size and capacity of the AAD system as well as the supporting NARANET backbone. Another factor that needs to be considered is the implicit requirement to provide quality of service and acceptable levels of performance.

For Phase 3A, AAD services will be provided to the public on a 24/7 basis via the Internet. During Phase 3A, NARA will need to provide some off-hours operations and maintenance support that will impact staffing levels and costs. Insofar as the projected public demand is concerned, the requirements for the number of simultaneous users, network support, storage capacity, and performance in the Phase 2 and Phase 3A Requirements Documents will be estimates that will be subject to revision based upon experience. With the uncertainty in projected public demand, the real requirement is assumed to be that the AAD system must be designed to be scalable, user-friendly, and efficient with the potential to normally run in a “lights-out” mode of operations. The reader is referred to the Performance Monitoring and O&M Analysis of Alternatives white papers for additional details.

- **Archival Records Available to the Public.** During Phases 3A and 3B, the only records that are currently planned to be accessible to the public through AAD services on the Internet will be archival data files without access restrictions. Records that are restricted because they contain information that is national security classified or information that is otherwise restricted will not be accessible to the public through the AAD Project. Public use versions of otherwise restricted datasets (appropriately redacted) may be made accessible through AAD services during Phase 3A.
- **Restricted Records Access.** After the Phase 3A system is delivered and fully operational, NARA will need to assess whether it wants to use URTS to materialize, search, and retrieve classified and sensitive electronic information or whether it will

need to develop classified and sensitive versions of AAD to address this requirement. Should AAD be used, access to classified and sensitive information may be provided to appropriately cleared researchers with a need-to-know. Such access will be very strictly controlled within security enclaves, and the current plan is that such access will be physically limited to secure spaces in NARA facilities where a stand-alone, “air-gapped” AAD system is available. At the current time, should NARA develop classified and sensitive version of AAD it will require three, as follows:

- Unclassified, Unrestricted
  - Restricted and/or Sensitive, but Unclassified (SBU)
  - National Security - Top Secret/Sensitive Compartmented Intelligence (TS/SCI)
- **State Department Cable Traffic.** The State Department cable traffic and withdrawal cards are an important set of archival electronic records planned to be searchable and accessible within the AAD environment. An initial capability to search and retrieve a small number of sample cable messages was prototyped and demonstrated as part of Phase 1 of the AAD Project. This basic Phase 1 capability to search and retrieve State Department cable traffic and withdrawal cards needs to be properly considered, continued, and enhanced and part of the Phase 3A design and development effort. The Department of State will transfer the first major batch of its electronic cables from 1973 in June 2002.
- **IT Security.** In the development of the AAD services production capability, the AAD Project shall be developed to be compliant with a formal IT security architecture to provide robust security services in such areas as: confidentiality, user authentication, access controls, data integrity, assured service availability, non-repudiation, and audit services. The reader is referred to the Security Analysis of Alternatives white papers for additional discussion on the concepts for achieving security and quality of service.
- **Integration with NARANET.** In Phases 3A and 3B, it is currently anticipated that AAD services will be accessible to the public “at large” via an Application Service Provider (ASP) as discussed in the O&M Analysis of Alternatives white paper. Making AAD services available to the public via the Internet is currently anticipated to have a significant impact on network bandwidth and resource allocation for the AAD servers, whether on NARANET or at an ASP site. Accordingly, the AAD Project must work closely with the NARA Architecture Team to identify any problems and to plan accordingly. Also, the potential impact on performance and quality of service must be assessed via rigorous stress testing in a testbed environment prior to Phase 3A AAD services being made available to the public “at large” via the Internet.
- **Use of Thin Web Browser Clients and Formatting Results for Printing.** As a result of the Phase 1 and Phase 2 development, the AAD systems architecture anticipates that users will access the system via commercially available Web browser

clients including Internet Explorer and Netscape. The Phase 1 prototype and the Phase 2 pilot have demonstrated the capability to format query results (e.g., records) in a Web page for electronic delivery and display on a computer screen. Such formatting employs the use of horizontal and vertical scroll bars that are particularly appropriate to displaying and rendering large numbers of agency database records that may have many columns. The Web-based interface will allow users to search the attributes of electronic database collections made available by NARA. The interface will also allow users to formulate and submit SQL queries to electronic databases in a user-friendly manner.

Within the AAD system, user-friendly submission of the queries and the interpretation of the results will be facilitated by automatic (and user-selectable, optional) substitution of data element names and coded values contained in the “completed” metadata. Human-readable data element names may be contained and defined in the data layout. Coded values in the data itself will be defined in the code lists. In using the AAD services, it is implicitly understood that substitution of human-readable and human-interpretable terms for obscure data element names and coded data values does not otherwise eliminate the need for user access to documentation.

Within the AAD environment, the formatting of AAD query results for printing purposes will be largely constrained to the intrinsic capabilities of the COTS Web browsers and the users’ printer(s). Many of the electronic records contain large numbers of columns and rows that cannot be easily formatted to be printed to a fixed medium such as paper. This problem is analogous to the common problem of trying to format very large spreadsheets for printing purposes. The potential use of report generators and various electronic style sheet mechanisms for rendering and formatting query results will be an engineering design matter as addressed in the Formatting and Printing Analysis of Alternatives white paper.

- **Section 508 Compliance.** In Phases 3A and 3B, the AAD system shall comply with Section 508 requirements. In 36 CFR Part 1194, the Architectural and Transportation Barriers Compliance Board (Access Board) issued final functional and performance standards to ensure accessibility for electronic and information technology covered by Section 508 of the Rehabilitation Act Amendments of 1998. The rules became effective on February 20, 2001. Subject to further engineering design and development, the AAD system is assumed to have to comply with the Section 508 rules. A preliminary analysis of the rules indicates that the requirement for compliance will not need to be extended to facets of the project associated with development and maintenance of the system. The high-level requirements for Section 508 compliance have been addressed in a separate Section 508 Analysis of Alternatives white paper.
- **User Profiles.** It is currently envisaged that the AAD system will have three major “modes” of operation during Phase 2 and Phases 3A and 3B, each offering specific



users access to, and control of, appropriate data and functions. At a high level, user profiles will be developed to provide functional support to:

- **Public Users** – This profile will define AAD system functions that will be provided to, allowed, or permitted for the general public user in accordance with this Concept of Operations document. The public user in Phase 2 is assumed to be a registered researcher physically located at a NARA facility. The public user in Phases 3A and 3B is a person accessing the system via the Internet.
- **NARA AAD Staff Users** – This profile will define AAD system functions that will be provided to, allowed, or permitted for NARA staff. Functionality to support NARA AAD staff users will be further subdivided depending upon the specific function they are performing or are assigned at any given time (e.g., reference support, research support, completing metadata, reviewing/approving metadata, making descriptive entries, etc.). In assigning AAD functions to NARA management staff, it is anticipated that most of the functions of an operations and maintenance nature will be performed within NWME. Some specialized AAD functionality (under strict user profile controls) may also be assigned or allocated to other NARA offices (e.g., NWC, NWCC, NWCT, and NL) particularly to provide AAD services for non-accessioned electronic records.
- **AAD System Administrator.** This user profile will define AAD system functions that are specifically reserved for the AAD System Administrator. The System Administrator profile will provide complete and secure control over AAD hardware, software, data, and network resources. The System Administrator will be responsible for the day-to-day operations, user profile management, systems maintenance, administration, backup and recovery, performance monitoring, and overall control of the AAD system. In Phase 3A, the AAD System Administrator will be outsourced to an Application Service Provider (ASP). While the System Administrator is provided with full access to, and operational control of, all of the functionality within the AAD system, it must be emphasized that the System Administrator will strictly serve an administrative role, and not perform archival staff functions.
- **Support for Future Development via Implementation of Decision Memoranda.** Subject to further AAD engineering development and analysis of alternatives, it is anticipated that users will be able to conduct complex queries in a user-friendly manner that are beyond the scope and capabilities of the current metadata model contained within AERIC. Within the AAD Project, requirements for this capability are currently being addressed via a Decision Memorandum process. Wherever possible, simple workarounds are being developed. Examples of complex search and retrieval situations include: searching across multiple relational tables, handling of parent-child relationships, grouped codes, voluminous code lists, conditional codes,

proper handling of date/time fields, proper handling of geospatial data fields, non-normalized data fields, files with delimiters, etc. The reader is referred to the Complex Queries metadata white paper for detailed discussion of the key operational issues.

In designing and developing the Phase 2 and Phase 3 systems, it is anticipated that the capability to handle such complexities in the AAD system will be incrementally added and configuration managed and controlled as the project progresses. It is further anticipated that such incremental improvements will be incorporated as Version 2.X or Version 3.X releases of the AAD system.

#### **4.5 Important High Level Phase 3A Test and Evaluation Factors**

The following summarizes the major test and evaluation factors that are currently being addressed:

- User-friendliness, human factors, and ease-of-use in searching structured databases
- Adequacy, clarity, and usefulness to users of metadata and federal agency- and NARA-produced documentation (potentially delivered via on-line means)
- Effectiveness of links from NARA descriptions of electronic records; i.e., When researchers get access to the databases, do they find what they had come to expect based on the descriptions?
- The effectiveness of improved workflow associated with the metadata completion, review, and approval process
- Interface designs with ARC
- The effectiveness of the Records Description Tool (RDT)
- Response times and benchmarks for search and retrieval under certain scenarios and the need to establish quality of service (QoS) performance measures
- Re-use of AAD Project architectural components and future scalability of the selected approach
- Identification of technical solutions and design requirements to address the Decision Memoranda being developed during the metadata completion process.
- The effectiveness of security measures to ensure that AAD data integrity and assurances of service availability.
- The effectiveness of measures to ensure that AAD is as fully Section 508 compliant as is practical.
- Stress testing to ensure scalability

## **5 Summary of Impact**

The operational scenarios in Chapter 4 on this document describe the AAD system from a customer's point-of-view. This chapter discusses the impacts of the AAD Project from a NARA operational and organizational perspective.

### **5.1 Phase 1 Impacts**

The AAD Project has developed a working prototype that is currently being used to evaluate the feasibility and desirability of the concepts set forth in the Concept of Operations.

#### **5.1.1 NWME**

Preparation of Archival Series. Version 1.2 of this ConOps document mentioned that additional impact would be felt with the need to complete the AERIC metadata for high-interest databases and to gather documentation to be scanned and converted to an electronic form for demonstration purposes. NWME has experienced a considerable operational impact in the area of identifying and preparing datasets for inclusion into AAD. This effort is far beyond what they had initially anticipated. Efforts are underway to streamline and automate the workflow associated with the metadata completion process and the data copying and data loading process, nonetheless, given the nature of the historical data and documentation, this will continue to present a workforce challenge to NWME.

Data Migration. Version 1.2 of the ConOps document also anticipated that effort would also be needed to assist with transfer of appropriate APS catalog data, the migration of the *Title List* to ARC, and the transfer of selected data files to the AAD platform.

Administrative. NWME is also impacted by having to devote time to review and provide comments on AAD deliverables.

#### **5.1.2 NH, NPOL**

The greatest impact on NH and NPOL has been the time they have had to devote to reviewing AAD deliverables and providing comments. NPOL has also had to devote time to working with the AAD Team in developing the AAD/ARC interface requirements.

### **5.2 Phase 2 Impacts**

Phase 2 offered the first production capability. Yet, it will be very important across NARA to realize that the Phase 2 system is only a "pilot" system, designed to allow NARA to further refine the product before it is released to the public in Phase 3A.

### 5.2.1 NWME

Preparation of Archival Series. During Phase 2, NWME continued to experience significant impacts on its staff. This has included such activities as reviewing NARA and agency documentation concerning candidate series, reviewing the metadata, copying files, verifying the metadata completion work, and composing ARC descriptive information. This is all necessary because if the metadata is not user-friendly and intuitive, NARA's image will suffer and NWME will receive large numbers of requests for assistance. NWME has responded by formalizing its review and approval processes, but the increased workload continues to be a significant factor for NWME.

Reference Responsibilities. Another factor NW has considered as the AAD Project moved through Phase 2 is that NW will need to consider how it will assign reference responsibilities. For example, NW has held discussions to determine which units will handle reference for series that are both textual and at some point become electronic, such as the State Department cables. In other cases NW developed procedures to determine who will be responsible for performing reference services on datafiles that are accessioned finding aids to hard copy records.

Reference Services. NWME has used AAD to respond to researchers' requests for information related to the series that are part of AAD in Phase 2. With the Phase 2 the system being a pilot implementation, NWME staff has participated in numerous sessions designed to determine how to enhance the system's usability.

Administrative. NWME has continued to be impacted by having to review and prepare comments on a rather large number of deliverables.

### 5.2.2 NWC

Reference Responsibilities. During Phase 2, NWC staff, who have responsibility for reference in series that are impacted by AAD, have been involved in discussions on how roles and responsibilities will change.

### 5.2.3 NH, NPOL

Originally, it was thought that there would be significant impact on NH engineers and architects to ensure that the security architecture is in place, that NARANET has adequate bandwidth, and that there are sufficient client and server machines. Since Phase 2 AAD services have been provided via an Application Service Provider (ASP), most of these concerns have been mitigated. A Service Level Agreement (SLA) to ensure access to the AAD system and quality of service has been developed and implemented. During Phase 2, NH, NW, and NPOL personnel were also involved with design and architectural issues associated with the AAD services; the interface with ARC, APS, and the Web; and planning, programming, and budgeting for the anticipated rollout of the AAD services to the public via the Internet.

Administrative. NH and NPOL will continue to be impacted by having to review and prepare comments on a rather large number of deliverables.

### **5.3 Phase 3A and 3B Impacts**

The major Phase 3A and 3B operational impact will be that public users will be empowered to directly search and retrieve accessioned electronic records via the Internet.

#### **5.3.1 NWME**

Preparation of Archival Series. As with Phase 2, it is anticipated that shifts in responsibility and additional positions will be required in NWME to effectively groom the metadata, and to load and verify data within the system. This includes such activities as reviewing NARA and agency documentation concerning candidate series, reviewing the metadata, copying files, verifying the metadata completion work, and composing ARC descriptive information. This is all necessary because if the metadata is not user-friendly and intuitive, NARA's image will suffer and NWME will receive large numbers of requests for assistance. Policies and procedures for coping with this possibility will need to be developed. Shifts in responsibilities and additional staff may be needed to fix and triage the metadata and to provide the necessary consultation and outreach services to the public.

Reference Services. NWME staff will be required to provide additional support and help services. Additional NWME staff will likely be required to accommodate the anticipated public demand and growth. From an operational perspective, NARA may need to limit the scope of the search and retrieval services that are offered to ensure quality of service and adequate systems performance. From an operational perspective, NWME will also be effectively "open for business" on a 24/7 schedule. This will pose a significant operational impact.

Administrative. NWME will continue to be impacted by having to review and prepare comments on a rather large number of deliverables.

#### **5.3.2 NH, NPOL**

- If NARA chooses to host the application, NARANET with its gateway to the Internet will probably require additional bandwidth.
- There will likely be a requirement for additional personnel with a background in operations and maintenance (O&M) of large database systems to support the anticipated growth in public demand. At the current time, a 1.5 FTE level of effort is anticipated for the operations and maintenance of the AAD system during Phase 3. The reader is referred to the O&M Analysis of Alternatives white paper for a discussion of the major factors and potential alternatives to achieve this objective. With the potential for a significant impact on networks, additional design and development work is needed to determine whether an ASP approach will be viable for Phases 3A and 3B.

- In Phases 3A and 3B, operations will be 24/7 thus implying the responsibility for providing around-the-clock support with additional positions being required, unless a “lights-out” mode of operations can be developed.
- Adequate on-line and near-line storage resources will be required and a caching strategy will need to be implemented to serve high-demand and frequently requested data.
- The security architecture must be fully implemented and from an operational perspective, information security must be continually monitored. Note: while security needs to be considered and monitored in all project phases, the need for new security services will be most pressing when the AAD services transition to the Internet (with public access in Phase 3) from the Intranet environment in Phase 2.
- At the completion of Phase 3B the interface between ARC and AAD will have been defined and fully implemented, and NWME will have the ability to create and edit its ARC entries. AAD services will then interface with ARC.

Administrative. NH and NPOL will continue to be impacted by having to review and prepare comments on a rather large number of deliverables. Additionally, NPOL and Life-Cycle Staff (LCS) will be heavily involved in reviewing and approving descriptive data entered via the AAD Records Description Tool (RDT).

## List of Acronyms

AAD	Access to Archival Databases
ADAM	Archival Data Access Mechanism
AERIC	Archival Electronic Records Inspection and Control
API	Application Programming Interface
APP	Annual Performance Plan
APS	Archival Preservation System
ARC	Archival Research Catalog
ASCII	American Standard Code for Information Interchange
CD	Compact Disk
CD-R	Compact Disk – Recordable
CD-ROM	Compact Disk – Read Only Memory
CFR	Code of Federal Regulations
ConOps	Concept of Operations
COTS	Commercial-off-the-shelf
DOT	Department of Transportation
ERA	Electronic Records Archives
FOIA	Freedom of Information Act
FTP	File Transfer Protocol
FY	Fiscal Year
HTML	HyperText Markup Language
IT	Information Technology
ITA	Information Technology Architecture
ITMRA	Information Technology Management Reform Act
LAN	Local Area Network
MED	Managed Evolutionary Development
MS	Microsoft
N	Archivist of the United States
NA	Office of Administrative Services
NAIL	NARA Archival Information Locator
NARA	National Archives and Records Administration
ND	Deputy Archivist
NF	Office of the Federal Register
NGC	Office of the General Counsel
NH	Office of Human Resources and Information Services

NHP	Information Resources Policy and Projects Division
NHPRC	National Historical Publications and Records Commission
NHT	Information Technology Services Division
NIH	National Institutes of Health
NL	Office of Presidential Libraries
NPOL	Policy and Planning Staff
NR	Office of Regional Records Services
NW	Office of Records Services – Washington, DC
NWC	Access Programs
NWL	Center for Legislative Archives
NWM	Modern Records Programs
NWME	Electronic and Special Media Records Services Division
O&M	Operations and Maintenance
OCR	Optical Character Recognition
OFAS	Order Fulfillment and Accounting System
OMB	Office of Management and Budget
PKI	Public Key Infrastructure
QoS	Quality of Service
RG	Record Group
SAIC	Science Applications International Corporation
SQL	Structured Query Language
TOMP	Task Order Management Plan
TSG	Technical Standard and Guideline
WAN	Wide Area Network
WWW	World Wide Web
XML	Extensible Markup Language



## Appendix B

### Selected Bibliography

1. National Archives and Records Administration, *Basic Laws and Authorities of the National Archives and Records Administration*, available from NWME staff, undated.
2. National Archives and Records Administration, *Electronic Records Archives (ERA) Program*, NARA NOTICE 2000-074, January 19, 2000.
3. National Archives and Records Administration, *NARA Fiscal Year 2002 Annual Performance Plan*, April 9, 2001, accessible at <http://www.nara.gov/nara/vision/2002plan.html>.
4. National Archives and Records Administration, *Ready Access to Essential Evidence, The Strategic Plan of the National Archives and Records Administration, 1997 – 2007*, Revised 2000, accessible at <http://www.nara.gov/nara/vision/nara2000.html>.
5. National Archives and Records Administration, *Regulations for the Public Use of Records in the National Archives*, NARA General Information Leaflet Number 2, 1991.
6. SAIC, *Task Order Management Plan for Engineering Support Services and Prototyping Support to the National Archives and Records Administration (NARA) for the Archival Data Access Mechanism (ADAM) – NARA IT Project 99-04*, Version 1.0 (final), August 11, 1999.
7. *Concept of Operations (ConOps) Document for the Access to Archival Databases (AAD) Project*, Version 1.2, March 20, 2000.
8. *Requirements Document for the Access to Archival Databases (AAD) Project, Phase 2*, Version 4.1 (final), February 26, 2001.
9. *Requirements Document for the Access to Archival Databases (AAD) Project, Phase 3*, Version 4.1 (final), February 26, 2001.
10. *System Description and Data Model for the Access to Archival Databases (AAD) Project Phase 1 Prototype*, Version 1.0, July 16, 2001.
11. *Metadata Quality Assurance (QA) Test Script for the Access to Archival Databases (AAD) Project*, Version 2.0 (final), April 4, 2001.
12. *Phase 1 Prototype Compliance Test Script for the Access to Archival Databases (AAD) Project*, Version 1.0, March 28, 2001.
13. *Task Order Management Plan (TOMP) for Access to Archival Databases (AAD) Population of AERIC Metadata*, Version 2.2 (final), April 26, 2001.

14. NARA, AAD Project Services Specified in Enclosure 2, Statement of Work, under Request for Quotation (RFQ) NAMA-01-Q-0008, AAD Project Phase 2 and Phase 3 Systems Engineering, March 12, 2001.
15. SAIC Oral Presentation, *Access to Archival Databases (AAD) Project Services* in Response to RFQ NAMA-01-Q-0008, April 23, 2001.
16. NARA IT Architecture Documents:
  - Baseline Characterization Document
  - Systems Development Guidelines (with document templates)
  - Managed Evolutionary Development (MED) Process Model
  - NARA Web Architecture Document
17. *Analysis of Alternatives White Papers* prepared during Phase 1:
  - Alternatives for Handling of Agency Documentation in an Online Environment (Scanning and Conversion), May 25, 2001.
  - Re-engineering of the Title List to Facilitate NARA 1301 Compliance, May 25, 2001.
  - Testbed Alternatives, May 25, 2001.
  - Alternatives for Security, May 25, 2001.
  - Alternatives for Development of a Common, Enterprise-Wide, Metadata Server for NWME's Operations, May 25, 2001.
  - Methods to Identify Bottlenecks in Performance, to Tune the System, and to Throttle/Limit the Scope of Queries, May 25, 2001.
  - Alternatives for Formatting, Display, and Printing Results, May 25, 2001.
  - Alternatives for Handling Complex Queries with Metadata - The Challenges and the Limitations that Metadata Creates for AAD System Engineering, May 31, 2001.
  - Management Mechanisms for Operations and Maintenance of the AAD System (including Trouble Tickets and Service Level Agreement (SLA)), April 9, 2001.
  - Metrics and Benchmarks for Completion of the Metadata, May 25, 2001.
  - Alternatives for Section 508 Compliance, April 9, 2001.